# IT'S ABOUT TIME:

# Elementary Mathematical Reflections On the Special and General Theories of Relativity

**Volume 2: Mathematical Background**

By Roger Cooke

# Contents

# Preface

The present volume contains two kinds of material: (1) Topics in pure mathematics that offer additional insight into the text of Volume 1, such as Appendix 1, which is devoted to hyperbolic trigonometry and closely connected with the material in Chapter 1 of the first volume, and Appendix 6, which discusses what is meant by mathematical invariance and establishes the invariance of tensor quantities as a foundation for the use of coordinate systems in Chapters 5–7 of the first volume; (2) background material needed for comprehension of Volume 1, such as calculus of variations (Appendix 2), differential geometry (Appendix 4), and differential equations (Appendix 5). In addition, there were certain other topics that seemed like natural extensions of what was in the text, such as the Jacobi last-multiplier method and rigid-body mechanics (Appendix 5) and especially the point-set topological foundation for abstract manifolds (Appendix 3). I decided to put all these topics into appendices so as not to overburden the basic text of Volume 1 with too much material.

I doubt if anyone needs to read *all* of these appendices in order to understand the material in Volume 1, but no one can possibly take courses in all of these areas as an undergraduate. Consequently, many students will probably need to consult one or two of these appendices. My goal is to get a book that undergraduates can read to gain a basic understanding of relativity. The material in them is far less complicated than (for example) Chapter 6 of Volume 1, but taken together, they do amount to a considerable corpus of material.

The content of these appendices is as follows:

- Appendix 1 constitutes an exploration of some of the prominent features of hyperbolic plane trigonometry, using the relativistic composition of velocities as a model.
- Appendix 2 provides a derivation of the Euler equation in the calculus of variations, which is needed for the computation of geodesics. The climax of this appendix is the proof of Euler's elegant theorem that a particle moving over a surface without any tangential acceleration will move in a geodesic of the surface.
- Appendix 3 grew out of an "overflow" of material from Appendix 4. It is devoted to some useful parts of point-set topology that will put the topics developed in the book into better perspective. These are topics not assumed as background in Volume 1 and also not essential to understanding it. This information is really needed only for a thorough understanding of Appendix 4.
- Appendix 4 contains the essentials of the theory of manifolds, used in Chapters 5 and 6.

- Appendix 5 is an exposition of the basic theorem on the existence and uniqueness of the solution of an initial-value problem for a differential equation, analyticity of solutions, and methods of solution. It also devotes some space to a topic peripheral to the main portion of the text, namely the Jacobi last-multiplier principle. This topic seems to have disappeared from the standard curriculum in differential equations. Yet it was of historical importance and still gives insight into the differential equations of mathematical physics. As a method of solving differential equations, it is of limited use. But as a method of showing that the solution (even if you can't find it in closed form) is the integral of an algebraic function—what used to be called an *abelian integral*—it is of great interest.
- Appendix 6 is devoted to the general philosophical and mathematical question of when two mathematically defined objects are the same thing.

Roger Cooke
March 2016

## APPENDIX 1

# Hyperbolic Trigonometry

In the text of Chapter 1, we discovered the very useful transformation $U \mapsto u = c\tanh(U/k)$, whose inverse is $u \mapsto U = k\ln\left(\sqrt{(c+u)/(c-u)}\right)$. We observed that relativistic addition of collinear velocities $u \ +_L \ v = (u+v)/(1+uv/c^2)$ corresponds via this mapping to ordinary addition $U+V$. As a result one can learn hyperbolic geometry by simply transferring the rather easily derived formulas of relativistic velocity triangles to this new setting. The purpose of the present appendix is to derive some of the unique features of hyperbolic trigonometry from the corresponding facts in relativity theory. Among these features, which have no analogs in Euclidean or spherical geometry are the angle of parallelism, horocycles, and ultra-ideal points. These concepts will reveal the circularity in the attempted proof of the parallel postulate by ibn al-Haytham, which Omar Khayyam remarked on in the epigram at the beginning of Chapter 1.

### 1. The Angle of Parallelism

Consider a relativistic velocity triangle with one side $u$ and the angle $\eta = \pi/2$ held fixed, and let $v$ be the side opposite angle $\xi$, as shown in Fig. 1.

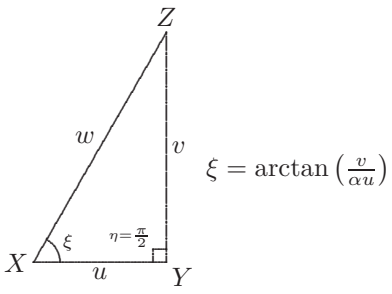


FIGURE 1. The tangent function in a right relativistic velocity triangle

According to our trigonometric formulas, we have

$$\cos\xi = \frac{u - v\cos\eta}{\sqrt{u^2+v^2-2uv\cos\eta - u^2v^2\sin^2\eta/c^2}} = \frac{u}{\sqrt{u^2+v^2-u^2v^2/c^2}}$$

$$\sin\xi = \frac{v\sin\eta}{\alpha\sqrt{u^2+v^2-2uv\cos\eta - u^2v^2\sin^2\eta/c^2}} = \frac{v}{\alpha\sqrt{u^2+v^2-u^2v^2/c^2}},$$

where $\alpha = (1-u^2/c^2)^{-1/2}$.

Therefore,

$$\tan\xi = \frac{v}{\alpha u}\,.$$

When $u$ is fixed, the angle $\xi = \arctan\big(v/(\alpha u)\big)$ is an increasing function of $v$, equal to 0 when $v = 0$. But, since $v$ cannot increase beyond its limit of $c$, there is a maximum value that $\xi$ can have in any triangle with $u$ fixed and $\eta = \pi/2$. That angle is

$$\xi(u) = \arctan\Big(\frac{c}{\alpha u}\Big)\,.$$

If we transfer this result to the corresponding hyperbolic triangle, where $u = c\tanh(U/k)$, $k$ being the unit of length, we find that

$$\xi(U) = \arctan\Big(\operatorname{csch}\frac{U}{k}\Big) = \operatorname{arccot}\Big(\sinh\frac{U}{k}\Big)\,.$$

This formula is most often written as

$$\cot\xi = \sinh\Big(\frac{U}{k}\Big)\,.$$

The limiting case occurs as $v \to c$, which is to say, $V \to \infty$, and in the limit we also have

$$w = \sqrt{u^2 + c^2 - u^2c^2/c^2} = c\,,$$

which corresponds to $W = \infty$. That is, the relativistic velocity triangle corresponds to a hyperbolic triangle two of whose sides have become infinite: The rays that these sides lie along never meet, although the ray from $X$ making any smaller angle than $\xi(U)$, which is called the *angle of parallelism* for the distance $u$, will meet the ray from $Y$ perpendicular to $XY$. The particular ray making angle $\xi(U)$ is called the *boundary parallel* to the ray $YZ$. It is uniquely determined. Any ray making a larger acute angle than $\xi(U)$ will also not meet the ray $YZ$, but (as will be shown below) the two rays will have a common perpendicular. Such a ray is called a *hyperparallel*.[1]

The Euclidean plane can be completed by agreeing that two parallel lines meet in an *ideal point* at infinity, whereas in elliptic (spherical) geometry any two lines meet in a *finite* point. We now see that in hyperbolic geometry, every line contains *two* ideal points, one at each end, and through each point not on a line, there are two boundary parallels to the given line, one meeting it at each of its two ideal points. As for the hyperparallels, it becomes convenient to say that two lines with a common perpendicular meet in an *ultra-ideal point*, a concept that will be elaborated below. For the present, it suffices to say that the ultra-ideal point of intersection is naturally paired with the *line* that is perpendicular to both of them, and so lines in the plane stand in for the ultra-ideal points. Thus, the hyperbolic plane has many subtle and interesting properties not found in the Euclidean plane.

It is intuitively obvious that the angle of parallelism must decrease as $U$ increases, and the formula in fact shows that it tends to 0 as $U \to \infty$ (that is, $u \to c$). In the other direction, for $U = 0$ ($u = 0$), we have $\cot\xi = 0$, and so $\xi$ tends to a right angle as $U$ tends to zero.

**Remark 1.1.** The statement of the parallel postulate by Euclid was that the ray from $X$ toward $Z$ will meet the ray from $Y$ toward $Z$ if $\xi$ is less than a right angle. Thus, in Euclidean geometry, the angle of parallelism is a right angle at all distances.

[1] For consistency with the notation we have adopted, we really ought to call it an *ultra-parallel*.

Since angles are absolute and are measured the same way in any form of geometry, a right angle can be taken as a universal unit of angle measure. We can assign any positive numerical value we choose to that unit, but given that we wish to use the standard Maclaurin series for the trigonometric functions, it would be foolish to assign any number other than $\pi/2$ to this angle. That is what we do, whether we are working in Euclidean or non-Euclidean geometry.

To sum up, given a line $l$ in a hyperbolic plane of curvature $-k^2$ (radius of curvature $k\sqrt{-1}$) and a point $P$ not on $l$, say at distance $u$ corresponding to $U/k$ from $l$, there are two distinct lines through $P$ that are boundary parallels to $l$ in opposite directions, making the angle $\xi(U)$ at $P$ on opposite sides of the perpendicular from $P$ to $l$. These two lines have no common perpendicular with $l$. Any line through $P$ making an angle less than $\xi$ with the perpendicular from $P$ to $l$ on either side of it will intersect $l$. Any line through $P$ such that the two mutually supplementary angles it makes with the perpendicular are both larger than $\xi(U)$ will have a common perpendicular with $l$ at some point, as we shall now prove.

**1.1. Hyperparallels.** Certain parts of geometry are absolute, that is, common to both parabolic (Euclidean) and hyperbolic geometry. One is the notion of a right angle, that is, an angle congruent to its supplementary angle. Although elliptic geometry, which is almost identical to the geometry of a sphere,[2] diverges from what we shall call absolute geometry in that there is no ordering on a line, and a line does not separate the plane into disjoint half-planes, almost all of what Euclid proved in the first two dozen or so propositions of his first book remain true in all three varieties of geometry. In particular, given a line and a point not on the line, it is always possible to draw a line through the given point that meets the given line at a right angle. Also, if two sides or angles of a triangle are unequal, the larger side is always opposite the larger angle. For hyperbolic and parabolic geometry, where the sum of the angles of a triangle is either less than two right angles or equal to two right angles, this last-mentioned result implies that the hypotenuse is the longest side of a right triangle. We do not find it necessary to assume that fact in our model of hyperbolic geometry, however, since it follows from the Pythagorean theorem $\cosh(C/k) = \cosh(A/k)\cosh(B/k)$ and the facts that the hyperbolic cosine function is (1) larger than 1 for positive values of its argument and (2) an increasing function of its argument. [3]

One consequence of this principle is the following lemma, which is valid in absolute geometry, and in particular in hyperbolic geometry:

**Lemma 1.1.** *Let $\angle PLK$ be a right angle, and let $PM$ be a ray emanating from $P$ and making an acute angle with the line segment $PL$, as shown in Fig. 2, with $M$ and $K$ on the same side of the line through $P$ and $L$. Then there is a point $Q$ on the ray such that the line segment $QR$ from $Q$ perpendicular to the ray $LK$ is shorter than $PL$.*

[2] The thing that keeps it from being actually identical is that a pair of antipodal points on a sphere must be regarded as a single point.

[3] In elliptic geometry, where a triangle may have both a right angle and an obtuse angle, it is not true that the side opposite a right angle is the largest. It is true, however, if the other two angles are acute. That fact follows from the Pythagorean relation $\cos(C/r) = \cos(A/r)\cos(B/r)$ and the facts that the cosine function is (1) never larger than 1 and (2) a positive decreasing function on the interval $(0, \pi/2)$.

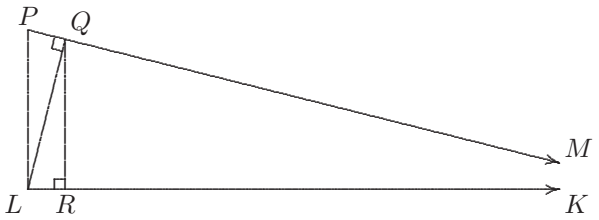


FIGURE 2. A ray $PM$ making an acute angle with the perpendicular $PL$ from $P$ to a given line $LK$. There are points like $Q$ on the ray that are closer to the line $LK$ than $P$ is ($PL > QR$).

PROOF. Let $Q$ be the foot of the perpendicular from $L$ to the ray $PM$ and $R$ the foot of the perpendicular from $Q$ to $LK$. Then $PL$ is the hypotenuse of triangle $PLQ$, and hence longer than $LQ$. But $LQ$ in turn is the hypotenuse of the triangle $LQR$, and hence longer than $QR$. Thus, $PL > QR$. □

From this lemma, we can see that if $PS$ is a boundary parallel ray to $LK$, then any ray emanating from $P$ on the same side as $PS$ and making a larger acute angle with $PL$ than $PS$ makes will have a common perpendicular with the given line. The situation is illustrated Fig. 3, where $PS$ is boundary parallel to ray $LK$, and angle $LPT$ is acute but larger than angle $QPS$. By the lemma, there are points on $PT$ closer to the ray $LK$ than $P$ is. Now if $W$ is chosen on $PS$, so that $PW$ is given as

$$PW = \frac{k \operatorname{arcsinh}\left(\tanh\left(PL/k\right)\right)}{\tan(\angle LPT - \angle LPS)},$$

the perpendicular to $PS$ at $W$ will meet the ray $PT$, in a point (which we shall assume without any loss of generality is $T$ itself) such that $TW$ is equal to $PL$. This follows easily from the trigonometric relations in a right hyperbolic triangle. Now let $X$ be the foot of the perpendicular from $T$ to the ray $LK$. Then $TX$ intersects the ray $PS$ in a point $Y$ since $PS$ lies between $T$ and $LK$. Then: (1) $TX$ is longer than $TY$, since $Y$ lies between $T$ and $X$; (2) $TY$, being the hypotenuse of the right triangle $TWY$, is longer than $TW$, and hence longer than $PL$; (3) by the lemma, there are points on the ray $PT$ that are closer to $LK$ than $P$ is. Therefore the minimum distance from a point on the line segment $PT$ to the ray $LK$ occurs at a point $U$ between $P$ and $T$. If $V$ is the foot of the perpendicular from $U$ to $LK$, then $UV$ is a common perpendicular of the two rays. (If $UV$ were not perpendicular to $PT$, then, by the lemma, there would be points of $PT$ near $U$ and closer to $LK$ than $U$ is, contradicting the minimality of the distance from $U$ to $LK$.)

The reader will find no difficulty in proving that in hyperbolic geometry, two lines can have at most one common perpendicular (Problem 1.3). Putting this fact another way, there are no rectangles in hyperbolic geometry.

In case the reader is wondering, a boundary parallel cannot be a hyperparallel. That is, two lines that are boundary parallels of each other cannot have a common perpendicular (Problem 1.5). Thus, we find the following trichotomy for a pair of lines $m$ and $n$ in a hyperbolic plane:

1. The lines may intersect in a point $P$ of the plane. (In elliptic geometry, this is the only possibility.)

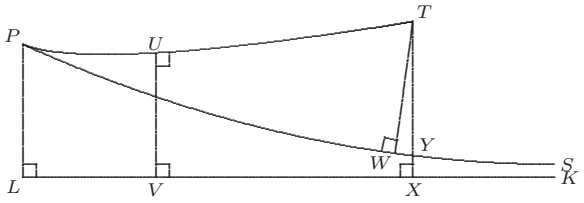


FIGURE 3. Lines with a common perpendicular. Because this situation is impossible in Euclidean geometry, it is necessary to draw some of the lines as if they were curved.

2. The lines may be boundary parallel in one of two directions, but not both. In that case, we say that the lines meet in an *ideal point*. Every line in the hyperbolic plane has two ideal "endpoints" at opposite ends, in which it meets its boundary parallels. This is in contrast to Euclidean/parabolic geometry, where parallelism is two-sided, and the two ideal points are fused into a single "point at infinity" in which all lines parallel to the line are said to meet.
3. The lines may have a common perpendicular. In that case, we say that the lines meet in an *ultra-ideal point*. We could, if we liked, call each line an ultra-ideal point and say that it is the point where two lines intersect if they are both perpendicular to that line. If we do that, then each finite point $P$ of a line $m$ can also be thought of as an ultra-ideal point, namely the ultra-ideal point identified with the line perpendicular to $m$ at $p$. In Euclidean/parabolic geometry, lines with a common perpendicular are simply the same thing as parallel lines, so that these last two cases fuse into one, and we have a dichotomy rather than a trichotomy. In the oddest case of all, elliptic geometry, even though any two lines intersect in a finite point and there are no ideal or ultra-ideal points, we find easily that two lines $m$ and $n$ always have a common perpendicular, namely the line determined by the point $P$ on $m$ farthest from $n$ and $Q$, the foot of the perpendicular from $P$ to $n$.

**1.2. The pseudo-sphere.** As we have remarked, the hyperbolic plane can be thought of as a sphere of imaginary radius, just as the elliptic plane can be thought of as an ordinary sphere of real positive radius, with antipodal points identified. We presume the ordinary sphere is sufficiently familiar not to require any "operating instructions," but the pseudo-sphere is not an object one normally sees or hears about. Accordingly, we shall devote this subsection to developing some of its properties.

On the infinitesimal level around each of its points, we can think of a sphere as having been obtained from a flat surface by first bending it slightly downward along one line to make a cylinder of radius, say $R$, then bending it downward to make a cylinder of radius $R$ along a line perpendicular to the first line, and then doing an identical bend along *every* line through the point. The result is a convex cap that forms a small neighborhood on the sphere. The curvature of the surface is defined to be the product of the curvatures of the circles formed on the first two bends, as discussed in detail in Chapter 5. Since that curvature is $1/R$ in both

cases, the curvature of a sphere is $1/R^2$, and its radius of curvature is $R$, the radius of the circle itself.

The pseudo-sphere is thought of similarly, except that after the first downward bend, making a circle of radius $k$, the second bend, along the perpendicular line, is *upward*, and then the amount of bending along each line making an angle between the two is obtained by interpolation of the amount of bending at the two extremes. In, particular along the bisector of the angle between the two directions, the amount of bending is zero. The result is a small piece of surface that resembles a saddle, in that the center point of it is a maximum as you travel in one direction and a minimum as you travel in the perpendicular direction, just as the center of a saddle is the highest point reached when drawing a line from one side of the horse to the other, but the lowest point reached when drawing a line from the front of the horse to the back. The curvature at the second bend is $-1/k$, since this is an upward bend. As a result, the curvature of the pseudo-sphere is $-1/k^2$ at every point. Its radius of curvature, which we call the radius of the pseudo-sphere, is $k\sqrt{-1}$.

To get a surface that has this saddle appearance at every point, one needs to revolve a graph that is convex on the side away from an axis about that axis. The position that the graph occupies at any given angle of rotation is a pseudo-longitude line or "circle," and the circle traced by each point of the graph is a pseudo-latitude circle. Not just any curve that is convex on the side away from the axis of rotation, however, will generate a surface with the same constant negative curvature at every point. By the rules used in differential geometry for computing curvature (see Chapter 5), it turns out that the only curve that will yield such a surface is the *tractrix*, whose equation in the $xz$-plane is[4]

$$\begin{aligned} z &= k\left(\operatorname{arcsech}\left(\frac{x}{k}\right) - \sqrt{1-\left(\frac{x}{k}\right)^2}\right) \\ &= k\left(\ln\left(\frac{k}{x}+\sqrt{\left(\frac{k}{x}\right)^2-1}\right) - \sqrt{1-\left(\frac{x}{k}\right)^2}\right). \end{aligned}$$

The tractrix is characterized as a curve whose subtangent (the portion of its tangent line between the point of tangency and the axis) is of constant length. It is shown in Fig. 4. Revolving it about the $z$-axis produces half of a pseudo-sphere, which we shall call a *pseudo-hemisphere*. The pseudo-hemisphere is shown in Fig. 5 below.

Because of the limited domains of the inverse hyperbolic secant and square root functions, there are points on this graph only for $0 < x \leq k$, and $z = 0$ when $x = k$. When the graph is revolved about the $z$-axis, the point $(k, 0)$ generates the equatorial circle of the pseudo-hemisphere. The complete pseudo-sphere is obtained by adjoining the reflection of this pseudo-hemisphere through the plane of the equatorial circle. The curvature of the resulting surface is not defined along the equatorial circle, since the surface has a "kink" there. Thus, our model is not a perfect embedding of the hyperbolic plane in $\mathbb{R}^3$. In a 1901 article in the *Transactions of the American Mathematical Society*, David Hilbert (1862–1943) showed that no such embedding exists.

Because the equation of a surface of revolution about the $z$-axis is independent of the angular coordinate, the equation of the pseudo-hemisphere in cylindrical

[4] Given that this surface is a model of part of the hyperbolic plane, the reader will not be surprised to find hyperbolic functions involved in its equation.

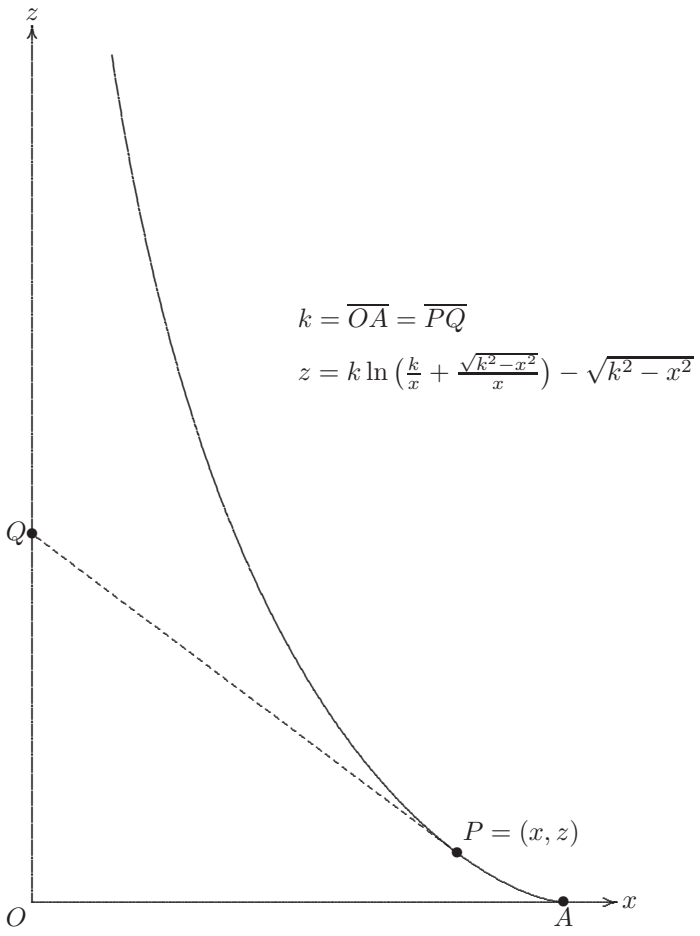


FIGURE 4. The tractrix in the $xz$-plane. Revolving it about the $z$-axis produces a pseudo-hemisphere.

coordinates $(r, \theta, z)$ is obtained by simply replacing $x$ with $r$:

$$\begin{aligned} z &= z(r,\theta) = k\left(\operatorname{arcsech}\left(\frac{r}{k}\right) - \sqrt{1-\left(\frac{r}{k}\right)^2}\right) \\ &= k\left(\ln\left(\frac{k}{r} + \sqrt{\left(\frac{k}{r}\right)^2 - 1}\right) - \sqrt{1-\left(\frac{r}{k}\right)^2}\right). \end{aligned}$$

The differential geometry of the pseudo-hemisphere is not complicated. When it is parameterized using $r$ and $\theta$ as independent parameters, its element of arc length is given by

$$ds^2 = dr^2 + r^2\,d\theta^2 + dz^2 = \left(1 + \left(\frac{dz}{dr}\right)^2\right) dr^2 + r^2\,d\theta^2 = \frac{k^2}{r^2}\,dr^2 + r^2\,d\theta^2\,.$$

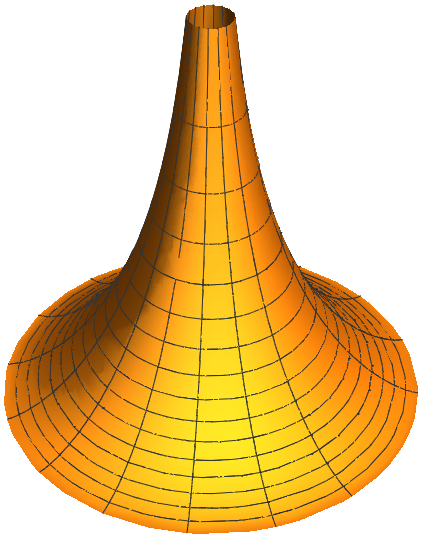

FIGURE 5. The pseudo-hemisphere

It is an easy exercise (Problem 1.6) to show that there are two kinds of geodesics[5] passing through a point $(r_0, \theta_0, z_0)$. One is a meridian of "longitude," which, when arc length on it is denoted $s$, has the parametrization

$$s \mapsto \left(r_0 e^{s/k}, \theta_0, z\big(r_0 e^{s/k}, \theta_0\big)\right), \quad -\infty < s < c \ln\left(\frac{k}{r_0}\right).$$

The actual length of the arc between $r_1$ and $r_2$ is $k\big|\ln\left(r_1/r_2\right)\big|$.

The other kind of geodesic is best expressed using $\theta$ as a parameter. It is conveniently indexed by the polar coordinates of the point where it crosses a meridian at right angles. If those coordinates are $(r_0, \theta_0)$, the parametrization is

$$\theta \mapsto \left(\frac{kr_0}{\sqrt{k^2 - r_0^2(\theta-\theta_0)^2}}, \theta, z\Big(\frac{kr_0}{\sqrt{k^2 - r_0^2(\theta-\theta_0)^2}}, \theta\Big)\right),$$
$$\theta_0 - \sqrt{\left(\frac{k}{r_0}\right)^2 - 1} < \theta < \theta_0 + \sqrt{\left(\frac{k}{r_0}\right)^2 - 1}\,.$$

(See Problems 4.5 and 4.6 in Volume 1 for the proof that these curves are geodesics.)

The second kind of geodesic, which we shall label $G(r_0, \theta_0)$ emerges from the equatorial circle, rises to a maximum height of $z(r_0, \theta_0)$ when $\theta = \theta_0$, and then descends to the equatorial circle again. The element of arc length along the geodesic

[5] On geodesics, see Appendix 2.

$G(r_0, \theta_0)$ is

$$ds = \frac{k^2 r_0}{k^2 - r_0^2(\theta - \theta_0)^2}\, d\theta\,,$$

and the length of the arc between parameter values $\theta_1$ and $\theta_2$ is

$$k\left|\operatorname{arctanh}\Big(\frac{r_0}{k}(\theta_2 - \theta_0)\Big) - \operatorname{arctanh}\Big(\frac{r_0}{k}(\theta_1 - \theta_0)\Big)\right| = \frac{k}{2}\left|\ln\left(\frac{\big(k + r_0(\theta_2 - \theta_0)\big)\big(k - r_0(\theta_1 - \theta_0)\big)}{\big(k - r_0(\theta_2 - \theta_0)\big)\big(k + r_0(\theta_1 - \theta_0)\big)}\right)\right|.$$

It is not difficult to prove that for any two distinct points on the pseudo-hemisphere, there is a unique geodesic joining them. If the points are $\big(r_1, \theta_1, z(r_1, \theta_1)\big)$ and $\big(r_2, \theta_2, z(r_2, \theta_2)\big)$ and $\theta_1 = \theta_2$, then the geodesic is of the first kind, and the distance between the two points is easily computed to be $|\ln(r_2) - \ln(r_1)|$. If $\theta_2 \neq \theta_1$, the geodesic joining the two points is $G(r_0, \theta_0)$, where

$$\begin{aligned} r_0 &= \frac{2k r_1^2 r_2^2 |\theta_2 - \theta_1|}{\sqrt{r_1^4 r_2^4 (\theta_2 - \theta_1)^4 + 2k^2 r_1^2 r_2^2 (r_1^2 + r_2^2)(\theta_2 - \theta_1)^2 + k^4 (r_2^2 - r_1^2)^2}}\,, \\ \theta_0 &= \frac{\theta_1 + \theta_2}{2} - \frac{(r_2^2 - r_1^2)k^2}{2 r_1^2 r_2^2 (\theta_2 - \theta_1)}\,. \end{aligned}$$

All of these results are a matter of routine algebra, involving nothing worse than solving quadratic equations. Here is a *Mathematica* program that, when given the parameter $k$, and the coordinates $(r_1, \theta_1)$, $(r_2, \theta_2)$ of the projections of two points $P_1$ and $P_2$ into the equatorial disk, will draw a pseudo-hemisphere and the geodesic joining $P_1$ and $P_2$. Like all computer programs, it has some limitations, since it cannot draw the entire pseudo-hemisphere. Some data will lead to a computation involving negative values of $r$, for example. The first line sets particular values of $k$, $r_1$, $r_2$, $\theta_1$, and $\theta_2$, which can be modified when the program is used interactively.

***Mathematica* Notebook 12.**

```
k = 10; r1 = 4k/5; θ1 = Pi/6; r2 = 3k/5; θ2 = 0;
ρ[r_, θ_] := {r Cos[θ], r Sin[θ],
k Log[(k + Sqrt[k^2 - r^2])/r] - Sqrt[k^2 - r^2]};
w[u_, v_, t_] := k u/Sqrt[k^2 - u^2 (t - v)^2];
θ3 = y /.  NSolve[{r1^2 (k^2 - x^2 (y - θ1)^2) ==
k^2 x^2, r2^2 (k^2 - x^2 (y - θ2)^2) == k^2 x^2}, {x, y}][[1]];
r3 = x /.  NSolve[{r1^2 (k^2 - x^2 (y - θ1)^2) ==
k^2 x^2, r2^2 (k^2 - x^2 (y - θ2)^2) == k^2 x^2}, {x, y}][[1]];
a = ParametricPlot3D[{r Cos[θ], r Sin[θ],
k Log[(k + Sqrt[k^2 - r^2])/r] - Sqrt[k^2 - r^2]},
{r, 1, k}, {θ, 0, 2 Pi},
AspectRatio -> Automatic, Axes -> False, Boxed -> False];
b = ParametricPlot3D[ρ[w[r3, θ3, t], t], {t, θ1,θ2},
AspectRatio -> Automatic, Axes -> False, Boxed -> False];
Show[a, b, PlotRange -> All]
```

**1.3. Euclidean representation of non-Euclidean planes.** Surfaces that we regard as flat obey the well-known rules of Euclidean geometry, at least within any conceivable limits of terrestrial measurement. This flatness of a sheet of paper

laid out on a table poses a problem for the representation of non-Euclidean spaces. Neither the elliptic plane, nor the hyperbolic plane can be faithfully represented in the Euclidean three-dimensional space $\mathbb{R}^3$. We can, however, represent convenient-sized pieces of these planes by a hemisphere and a pseudo-hemisphere respectively, with the equators removed. These hemispheres suffice to represent most finite figures such as triangles and quadrilaterals. The only drawback is that the surfaces are still embedded in three-dimensional space. To represent them as a figure in a book, we still need to project them onto a plane, and when we do that, lengths and angles still get distorted. Figures that are congruent do not *appear* to be so. A figure in a book, however, is generally used only as a *qualititative* guide to thought, so that a pedantic accuracy in drawing a figure is not necessary to convey the ideas one wishes to present. It is usually sufficient to draw a triangle with its sides bowed outward in elliptic geometry and bowed inward in hyperbolic geometry, as in Fig. 6. (See our discussion of Toponogov's theorem in Chapter 1.) That representation conveys the important information that the angle sum of a triangle is greater than two right angles in elliptic geometry and less than two right angles in hyperbolic geometry. An accurate drawing of two equilateral triangles, one on the sphere, and one on the pseudo-sphere, can be seen in Figs. 7 and 8. As you can see, the two-dimensional projection of these figures makes the sides and angles appear to be unequal. Thus, from now on, we shall draw figures that are only qualitatively, not quantitatively, accurate, as in fact we did in Fig. 3 above.

Even though we are forced to compromise in a flat-page representation, it should *not* be thought that the trigonometry we have discussed is a mere fiction. Both the hemisphere and the pseudo-hemisphere can be embedded isometrically in $\mathbb{R}^3$. It is easy enough to imagine living under a hemisphere, since most people have seen pictures of the inside of an igloo. It is a little trickier to imagine a tent shaped like a pseudo-hemisphere, but if you picture a teepee with sagging sides, you will have the idea. In any finite region of $\mathbb{R}^3$, the portion of the pseudo-hemisphere contained in that region must have a hole in the top.[6] If you looked at triangles on a hemisphere whose sides are drawn by stretching strings taut between the vertices—the strings would have to be stretched against the outside of the hemisphere—you really could verify the trigonometric relations of elliptic geometry. On the pseudo-hemisphere, one can find longitudinal geodesics this way, stretching strings against the *inside* of the surface; but a string stretched between two points not on the same longitudinal line of the pseudo-hemisphere would not stay taut against the surface on either side. (It would find a shorter path through $\mathbb{R}^3$ not lying on the surface.)

## 2. Circles, Horocycles, and Equidistant Lines

The last special feature of the hyperbolic plane that we wish to consider is the region between a circle and its tangent line. To that end, consider a line $n$ containing a point $T$ and a ray $TU$ from $T$ perpendicular to $n$, all shown in Fig. 9. Let $V$ be a point on $TU$, and draw the circle tangent to $m$ at $T$ and having center at $V$. Thus, we get a finite circle with radius $TV$. If the plane is the elliptic plane, you can use the hemispherical model for it, picturing $n$ as the equator and $TV$ as an arc of a line of longitude. The circle will then be a small circle (not a geodesic, that is, great circle) as long as $TV$ is less than the arc from the equator to the pole of the sphere. Once $V$ reaches the pole, the circle about $V$ through $T$ becomes identical

[6] Very convenient for venting the smoke from a campfire!

FIGURE 6. Flat-space depiction of triangles in elliptic (left) and hyperbolic (right) geometry

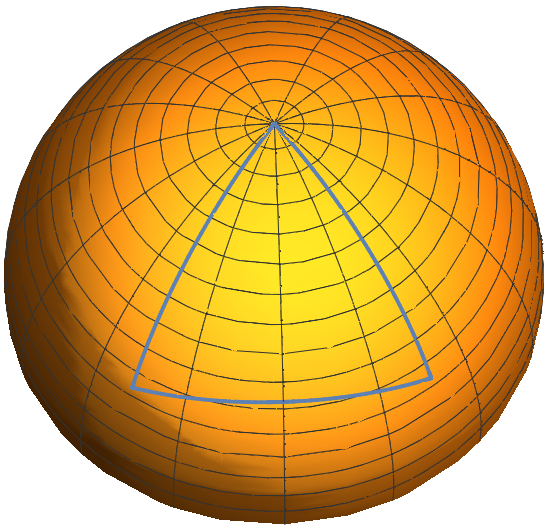

FIGURE 7. An equilateral triangle in a hemisphere

with the line $n$. Thus, lines in the elliptic plane are simply maximal circles (great circles), and there is nothing more to be said.

If the plane is the Euclidean plane, the point $V$ can recede all the way to infinity, producing larger and larger circles, which collectively engulf the entire half-plane on the side of $n$ containing $U$. The proof of that fact is that if $S$ is any point in the half-plane, then angle $UTS$ is less than a right angle, so that the perpendicular bisector of $TS$ intersects $TU$. (The ray $TU$ and this perpendicular bisector are cut by the transversal $TS$ and the sum of the interior angles is less than two right angles on the side containing $U$. By Euclid's form of the parallel postulate, this means that $TU$ and the perpendicular bisector meet in a point $V$. Since $V$ lies on

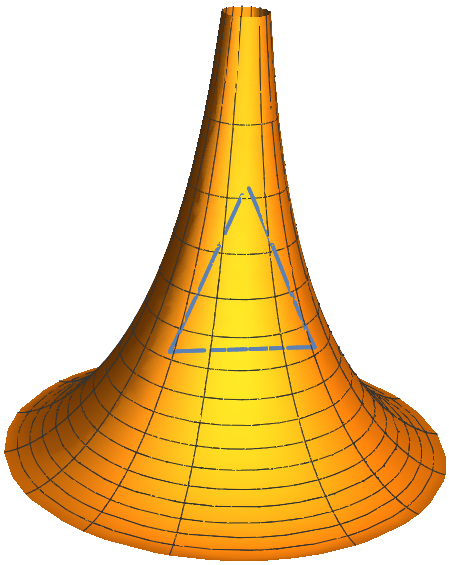

FIGURE 8. An equilateral triangle in a pseudo-hemisphere

the perpendicular bisector of $TS$, it is equidistant from $T$ and $S$. That is, both points $T$ and $S$ lie on the same circle about $V$.) There is no more to be said in the Euclidean case: Every point in the open half-plane on the side of $n$ containing $U$ lies on a circle with center at a finite point $V$ of the ray $TU$. In elliptic geometry, a line is a circle of finite (maximal) radius whose center is an actual point of the plane. In Euclidean geometry, a line is a circle of infinite radius whose center is an ideal point of the plane.

Finally, if the plane is the hyperbolic plane, and $TW$ is any ray in the half-plane on the side of $m$ containing $U$ and angle $UTW$ is acute but not zero, then there is some distance $d$ for which $\theta = \angle UTW$ is the angle of parallelism at distance $d$. If $Z$ is the point on the ray $TW$ for which the length of $TZ$ is $2d$, then the perpendicular bisector of $TZ$ is boundary parallel to $TU$. The boundary parallel through $Z$ to this perpendicular bisector also makes angle $\theta$ with $TZ$, and we now have three rays, all boundary parallels to one another. In that case, we say that $T$ and $Z$ lie on the same *horocycle* whose center is at the ideal point $\Omega$ common to the three boundary parallels. Thus, a horocycle is a circle of infinite radius whose center is an ideal point of the plane, but, in contrast to Euclidean geometry, it is not a line. One can show easily that all horocycles are congruent.

There still remain points of the half-plane that are not on any finite circle with center on the ray $TU$ and not on the horocycle with radius $T\Omega$. What can we say about them? If $W$ is any point in the region between the horocycle and the line $m$, then $TW$ intersects the horocycle in some point, so that the length of $TW$ is greater than twice the distance for which the angle $UTW$ is the angle of parallelism. Another way of saying this is that the angle $UTW$ is larger than the angle of parallelism at a distance equal to half the length of $TW$. But that means

that the perpendicular bisector of $TW$ has a common perpendicular $m$ with the line $TU$, and by the symmetry of the figure, $T$ and $W$ are equidistant from that common perpendicular. Thus $W$ lies on the locus of points equidistant from the line $m$. Taking $m$ to stand for an ultra-ideal point, we might say that $T$ and $W$ lie on the same "circle of ultra-infinite radius" with center at the ultra-ideal point $m$. What geometers actually do say, is that the two points lie on the same *equidistant curve* relative to $m$. The point we wish to emphasize is that in the hyperbolic plane an equidistant curve is *not* a straight line, as ibn al-Haytham (and myriad others since him) took for granted, thereby *assuming* a Euclidean plane.

The distinction between equidistant curves and parallel lines, the neglect of which has misled so many people like ibn al-Haytham,[7] who tried to prove the parallel postulate by defining parallel lines to be equidistant, can be understood by a thought experiment, in which a pair of railroad tracks is laid with one of the pair lying over the equator. That one is following a "straight line" (geodesic) on the surface of the earth. Its companion, which is equidistant from it at a fixed distance, is following a circle of latitude, which is not a geodesic. The effect is too small to be noticed and certainly plays no measurable role in the mechanics of the motion of any actual train. But it is present in theory, and has an important physical consequence: Since at least one of each pair of wheels on a train is not moving along a geodesic, it is subject to some small force, and therefore puts a (negligible) amount of stress on the axle. In general, the concept of a rigid body does not make sense in physics when non-Euclidean geometry is used, since only the points along one line of such a body can be moving along a geodesic at any given time; thus, motion of a solid body in non-Euclidean space is always accompanied by a strain, as Felix Klein remarked in his 1893 lectures on the spinning top at the World's Columbian Exposition (see the Preface to Volume 1). Particularly in special relativity, rigid bodies make no sense. If you had a rigid pencil for example, any word you wrote with the tip would be mirrored by the motion of the other end of the pencil, and thus information would be transmitted instantaneously, which is impossible.

The three cases are further illustrated in Fig. 10, where two rays emanating from points $P$ and $Q$ on the same side of the line through $P$ and $Q$ make equal acute angles with the line segment $PQ$. If those angles are less than the angle of parallelism for half the distance $PQ$, as illustrated by the angle $\theta_1$ on the left, then the rays meet at a finite point $O$ on the perpendicular bisector of $PQ$, and the curve $PRQ$ is an arc of a circle, which is perpendicular to every line through $O$. If the angles are exactly equal to the angle of parallelism for half the distance $PQ$, as illustrated by angle $\theta_2$ in the center, then both rays are boundary parallels of the perpendicular bisector of $PQ$ (and hence also boundary parallels to each other), and the curve $PRQ$, is part of the locus of a point that moves so as to remain always perpendicular to every line that is boundary parallel to these rays in the given direction. If we denote the ideal point on all these lines as $\Omega$, we can think of $\Omega$ as the center of an ideal circle, a *horocycle*.[8] Finally, if the acute angles are larger than the angle of parallelism, as illustrated by angle $\theta_3$ on the right, then the

[7] See Chapter 1 of Volume 1. Ibn al-Haytham's idea was re-invented in the twentieth century by Jeremiah J. Callahan, President of Duquesne University. See Chapter 8 of Volume 1.

[8] Of course, $\Omega$ lies an infinite distance away from $P$ and $Q$. This drawing is therefore not accurate. Nor is the drawing on the right, since (again) it represents a curve that does not exist in the Euclidean plane. The horocycle $PRQ$ in this figure and the equidistant curve $PRQ$ on the right

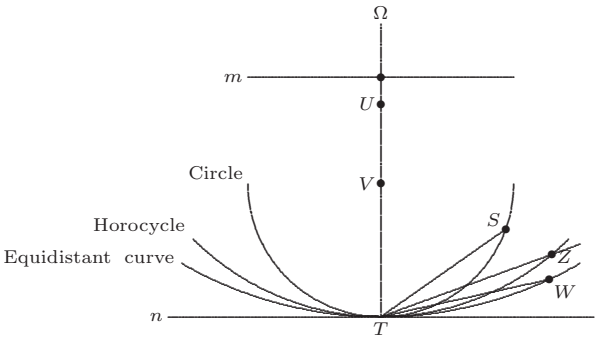


FIGURE 9. Angle $UTS$ is less than the angle of parallelism for half the distance $TS$, and $T$ and $S$ both lie on an arc of a finite circle with center at a point $V$ on the ray $TU$. Angle $UTZ$ is equal to the angle of parallelism for half the distance $TZ$, and $T$ and $Z$ both lie on an arc of the same horocycle with center at the ideal point $\Omega$ on the ray $TU$. Angle $UTW$ is acute, but larger than the angle of parallelism for half the distance $TW$, and $T$ and $W$ both lie on an arc of the same equidistant curve from a line $m$ perpendicular to $TU$.

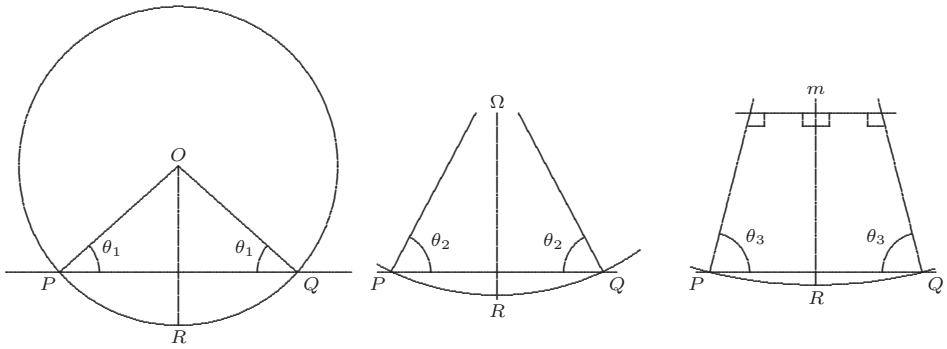


FIGURE 10. Left: Angle $\theta_1$ is less than the angle of parallelism for half the distance from $P$ to $Q$. Center: Angle $\theta_2$ is the angle of parallelism. Right: Angle $\theta_3$ is larger than the angle of parallelism, but still acute.

two rays and the perpendicular bisector of $PQ$ all have a common perpendicular $m$, which represents an ultra-ideal point that can be thought of as the center of an ultra-ideal circle. The distances from $P$ and $Q$ to $m$ are equal, and if we draw the entire locus of points lying at that distance from $m$ on the same side of $m$, part of that locus will be the curve $PRQ$ shown in the figure. The curve that we just referred to as an ultra-ideal circle is called an *equidistant curve*.

---

both collapse to become the line segment $PQ$ when the curvature of the hyperbolic plane becomes 0, which means that it degenerates to a Euclidean plane.

**Remark 1.2.** As a final comment on this mini-course in non-Euclidean geometry, we remark that if a horocycle is revolved about one of its radii, the result is a two-dimensional surface in three-dimensional hyperbolic space called a *horosphere*. Lobachevskii and Bólyai were able to work out the trigonometry of the hyperbolic plane by remarking that the geodesics on the horosphere are horocycles, and given a horocyle $h$ and a point $P$ on the horosphere not on $h$, there is one and only one horocycle on the horosphere passing through $P$ and not intersecting $h$. In short, the geometry of the horosphere is Euclidean geometry.

## 3. The Beltrami–Klein Model

Although an isometric copy of the entire hyperbolic plane does not exist in $\mathbb{R}^3$, it is nevertheless possible to construct a model of this plane within the context of Euclidean geometry. Such a model was first proposed by Eugenio Beltrami (1835–1900) in two 1868 articles: "Saggio di interpretazione della geometria non-euclidea," (*Giornale de matematiche* **VI**, 285–315) and "Teoria fondamentale degli spazi di curvatura costante," (*Annali di matematiche*, ser. II, **2**, 232–255). Beltrami's ideas were based on projective geometry, which we have not space to discuss. The same projective ideas were taken up by Felix Klein in 1871, in his article "Ueber die sogenannte Nicht-euklidische Geometrie," (*Mathematische Annalen*, **4**, 573–625). We have already mentioned the modern version of this model, though not by name, in Notebook 9, where we referred to it as the disk of relativistic velocities.

The model is based on the simple fact that the set of possible velocities that can be measured in a fixed reference frame can be recorded as the points in the interior of a disk of radius $c$ in $\mathbb{R}^2$, the point with polar coordinates $(r, \theta)$ being thought of as motion at speed $r$ in the direction making angle $\theta$ with a fixed axis of reference (the $x$-axis). We consider a triangle with vertices at the origin and the points with coordinates $(r, \theta)$ and $(r + dr, \theta + d\theta)$, as shown in Fig. 11.

The relativistic law of cosines derived in Chapter 1 asserts that for a triangle with sides (velocities) $u$, $v$, and $w$ and angle $\eta$ opposite $w$,

$$w^2 = \frac{c^4u^2 - 2c^4uv\cos\eta + c^4v^2 - c^2u^2v^2\sin^2\eta}{(c^2 - uv\cos\eta)^2}\,.$$

Taking $u = r$, $v = r + dr$, $\eta = d\theta$, and $w = ds$ and retaining only the lowest-order terms in the Maclaurin expansion of $ds^2$ as a function of $dr$ and $d\theta$, with $r$ fixed, we find

$$ds^2 = \frac{c^4}{(c^2 - r^2)^2}\,dr^2 + \frac{r^2c^2}{c^2 - r^2}\,d\theta^2\,.$$

One can show (see Problem 1.11 below) that the geodesics in this metric are simply the chords. If we think of the endpoints of the chords (on the circle of radius $c$) as ideal points, we see that each chord contains two ideal points. The exterior of the disk then corresponds to ultra-ideal points, since two disjoint chords that do not meet on the boundary circle may, when extended, meet outside the circle. It is possible, though we shall not take the time to do it, to define the angle between two intersecting chords in such a way that the trigonometry of this space is precisely hyperbolic trigonometry. The angles so defined are not what appears to the eye. We shall limit our remarks in this direction by saying only that two chords intersect in a right angle if the extension of one of them passes through the point of intersection of the tangents at the endpoints of the other. (It is by no means

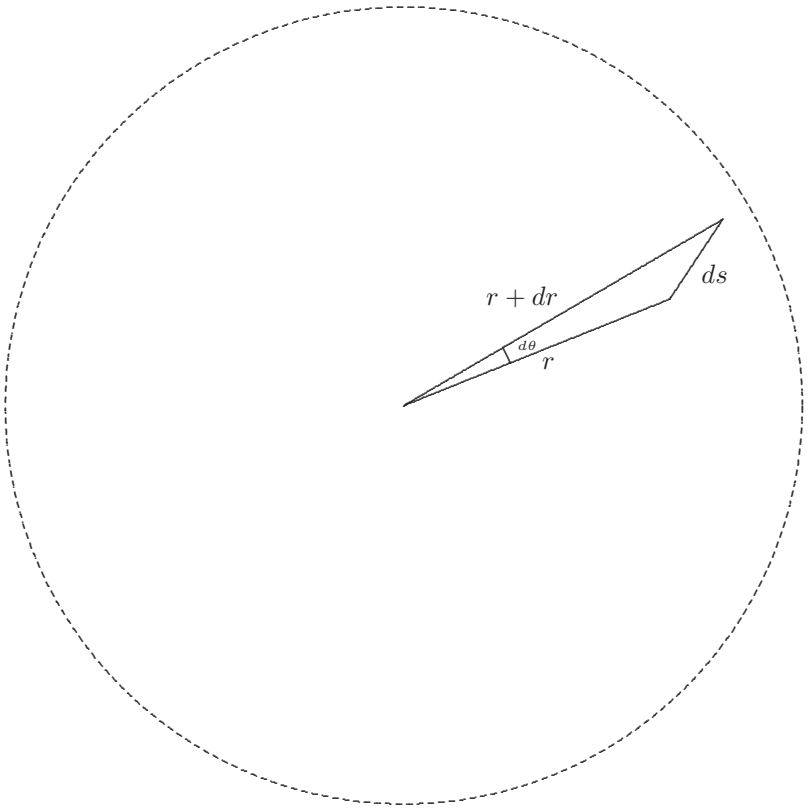


FIGURE 11. Metric $ds$ on a disk $c$ induced by the relativistic law of cosines.

obvious that this relation is symmetric between two chords.) In that case, one can see that two non-intersecting chords that do not share an ideal point will have a common perpendicular, as shown in Fig. 12.

**Remark 1.3.** Besides the Beltrami–Klein model of the hyperbolic plane, there is another that is commonly encountered, due to Henri Poincaré (1854–1912) and known consequently as the Poincaré disk model. Two kinds of curves stand in for lines in this model: (1) diameters of the disk, and (2) circular arcs that meet the boundary of the disk at a right angle, as shown in Fig. 13. Since this model does not follow naturally from the Lorentz transformation, we content ourselves with just this mention of it and will not develop its properties, remarking only that, in contrast to the Beltrami–Klein model, it does represent angles accurately.

The tiling of the hyperbolic plane in the Poincaré disk model has been illustrated in a number of paintings by Maurits Cornelis Escher (1898–1962). Escher apparently became acquainted with this model in a letter from the mathematician Harold Scott MacDonald Coxeter (1907–2003) of the University of Toronto, who sent him a sketch of the tiling in 1957. One of the best illustrations is the "angels and demons" sketch shown below and known prosaically as *Circle Limit IV* (one of many such sketches in this genre).

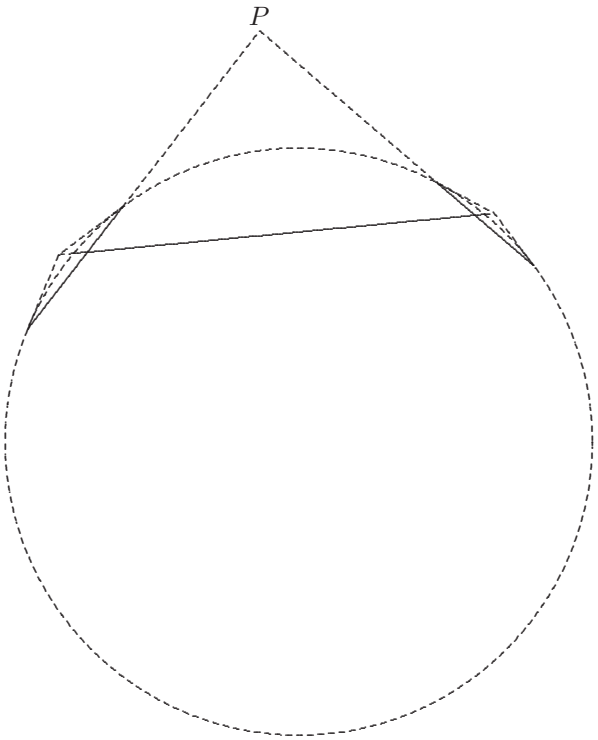


FIGURE 12. Two disjoint, non-boundary parallel lines in the Beltrami–Klein model, showing their common perpendicular and the ultra-ideal point $P$ in which they intersect.

Escher's *Circle Limit IV*, a tiling of the Poincaré disk model of the hyperbolic plane. Under the metric used in this disk, any two of the "angels" are congruent, as are any two of the "demons."

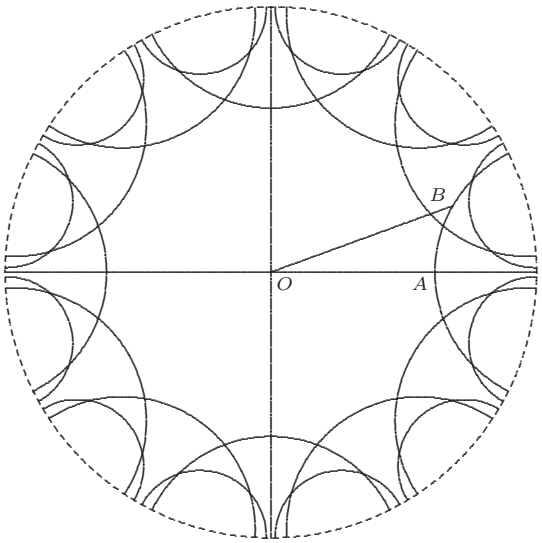


FIGURE 13. The Poincaré disk model of the hyperbolic plane showing a selection of geodesics, including the geodesic right triangle $OAB$.

## 4. Spherical Trigonometry

To nail down the last corner of our brief discussion of non-Euclidean geometry, in this section we shall derive the fundamental facts about spherical triangles, since we made use of them in Chapter 1. As far as triangles of modest size are concerned, an ordinary hemisphere provides the same perfect model for elliptic geometry that the pseudo-hemisphere provides for hyperbolic geometry. The trigonometric formulas are in strict analogy with those of hyperbolic geometry, with the hyperbolic functions cosh, sinh, and tanh being replaced by the standard trigonometric functions cos, sin, and tan, thereby illustrating Lambert's remark that hyperbolic geometry would be the geometry of a sphere of imaginary radius.

We assume the reader knows the Euclidean law of cosines, and we shall apply it to the faces of a tetrahedron to derive the corresponding law of cosines for a spherical triangle, that is, a triangle whose sides are arcs of three great circles on the sphere.

The situation is shown in Fig. 14. In that figure, the spherical triangle $ABC$ on a sphere with center at $O$ has sides $a$, $b$, and $c$ respectively opposite its vertices $A$, $B$, and $C$. The lengths of these arcs are measured relative to the radius $r$ of the sphere as a unit length. These ratios are the radian measure of the angles the corresponding arcs subtend at the center $O$. For example, $a/r$ is the radian measure of the angle $\angle BOC$.

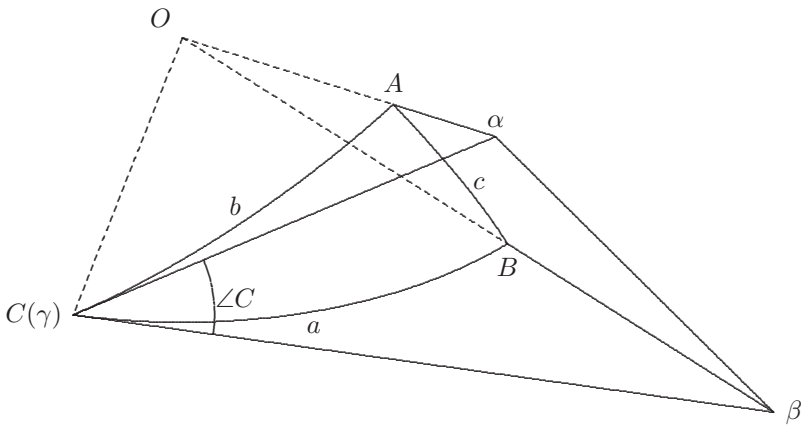


FIGURE 14. A spherical triangle ($ABC$) and its enclosing tetrahedron ($O\alpha\beta\gamma$)

As for the angles of the spherical triangle itself, the only one we wish to consider is the angle at vertex $C$, where the arcs of length $a$ and $b$ meet.[9] This angle also is expressed in radian measure. By definition, it is the angle between the tangents to sides $a$ and $b$ at $C$. These tangents meet the extensions of the radii through $A$ and $B$ in points $\alpha$ and $\beta$ respectively, forming a tetrahedron $O\alpha\beta\gamma$, where $\gamma$ coincides with $C$. It is best to visualize this figure by thinking of $O$, $B$, and $C$ (and hence also $\beta$ and $\gamma$) as being in a horizontal plane with $A$ and $\alpha$ above that plane, $\alpha$ being the highest point in the figure.

The secret of solving the spherical triangle, given that you know arcs $a$, and $b$ and the angle $C$ opposite arc $c$, is to solve all four triangular faces of the enclosing tetrahedron. These four faces are shown in Fig. 15. The two right triangles on the left in that figure contain angles whose radian measures are $a/r$ and $b/r$. The leg adjacent to those angles, in both cases, is equal to the radius $r$ of the sphere. Therefore the hypotenuses are respectively $r\sec(a/r)$ and $r\sec(b/r)$ and the sides opposite are $r\tan(a/r)$ and $r\tan(b/r)$. These lengths can then be transferred to the two scalene triangles on the right, both of which contain the edge $\alpha\beta$, whose length we denote by $\ell$. By the law of cosines for plane triangles, we find

$$\begin{aligned}\big(r\tan(a/r)\big)^2 + \big(r\tan(b/r)\big)^2 - 2r^2\tan(a/r)\tan(b/r)\cos(\angle C) &= \ell^2 \\ = \big(r\sec(a/r)\big)^2 + \big(r\sec(b/r)\big)^2 - 2r^2\sec(a/r)\sec(b/r)\cos(c/r)&\,.\end{aligned}$$

Obviously, we can cancel $r^2$ from this equation. Also, since $\sec^2 x = 1 + \tan^2 x$, we can rewrite this equation as

$$-2\tan(a/r)\tan(b/r)\cos(\angle C) = 2 - 2\sec(a/r)\sec(b/r)\cos(c/r)\,.$$

[9] For later reference, we note that $\angle C$ is the *dihedral angle* between the planes of the triangles $OAC$ and $OBC$. By definition, the dihedral angle between two half-planes meeting in a line—in this case, the line $O\gamma$—is the angle between two rays emanating from the same point on the line of intersection, one in each half-plane, and each perpendicular to the line of intersection.

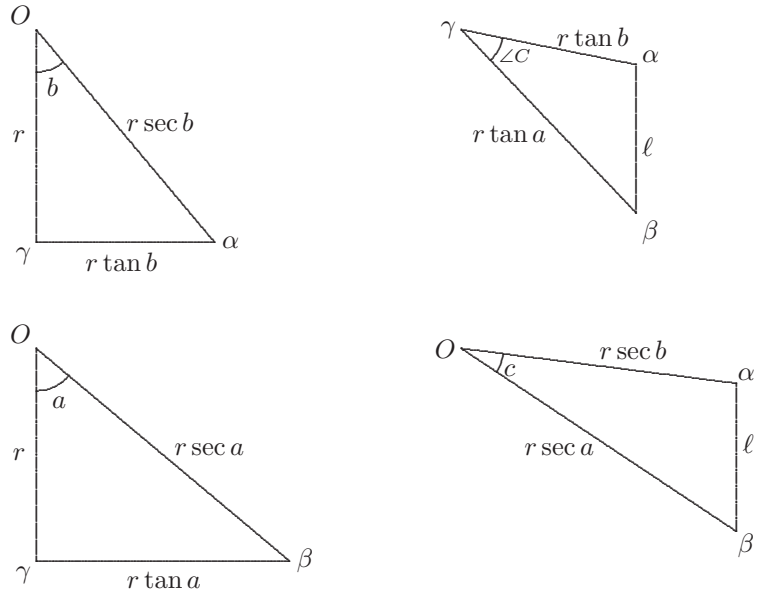


FIGURE 15. The enclosing tetrahedron cut open

If we multiply both sides by $-\frac{1}{2}\cos(a/r)\cos(b/r)$ and rearrange terms, we arrive at the spherical law of cosines:

$$\cos(c/r) = \cos(a/r)\cos(b/r) + \sin(a/r)\sin(b/r)\cos(\angle C)\,.$$

This formula can actually be used for navigation on the surface of the earth. But the main use we have for it is in Chapter 1, where it is used to compute the third angle at a trihedral vertex, given two of the angles, and the dihedral angle between the planes they lie in. In that form, it applies in any geometry whatever, not just Euclidean.

The special case of the law of cosines that occurs when $\angle C$ is a right angle is called the spherical Pythagorean theorem:

$$\cos(c/r) = \cos(a/r)\cos(b/r)\,.$$

**Remark 1.4.** Although we have stated the spherical law of cosines in terms of the arc lengths $a$, $b$, and $c$ of the sides of a spherical triangle, as

$$\cos(c/r) = \cos(a/r)\cos(b/r) + \mathit{sin}(a/r)\sin(b/r)\ \cos(C)\,,$$

where $C$ is the angle opposite the side of length $c$ and $r$ is the radius of the sphere, the formula was derived in Euclidean geometry, where the ratios $a/r$, $a/r$, and $c/r$ are just the radian measures of the angles the three sides subtend at the center of the sphere. The angle $C$ in turn is the dihedral angle between the planes $OAC$ and $OBC$, and can be denoted $\gamma$. If these ratios are replaced by those angles and denoted $A$, $B$ and $C$, we get a formula that applies to any three planes meeting at a point $O$ in any geometry whatsoever, since angles are absolute and common to all geometries, namely

$$\cos C = \cos A\,\cos B + \sin A\,\sin B\,\cos\gamma\,.$$

This is the form in which we used the theorem in Chapter 1 of the text.

## 5. Problems

**Problem 1.1.** The formula for the angle of parallelism provides a way of determining the curvature of the particular hyperbolic plane one happens to be on. Show that the distance $U$ for which the angle of parallelism is half of a right angle is

$$U = k \ln\left(\sqrt{2} + 1\right).$$

Thus, if (1) physical space is hyperbolic and (2) it actually is possible to determine the angle of parallelism experimentally—it probably isn't—this equation can be solved for the radius of curvature $k\sqrt{-1}$. What velocity does this natural unit of length correspond to when translated into the space of relativistic velocities?

**Problem 1.2.** In a manuscript that was not published during his lifetime, Gauss defined the area of a hyperbolic triangle to be its defect times $k^2$, where $k$ is the radius of curvature of hyperbolic space. He then gave a formula for the least upper bound of the areas of all finite triangles in terms of $C^2$, where $C$ is the absolute unit of length in Problem 1.1. What is that least upper bound?

**Problem 1.3.** Prove that two lines in the hyperbolic plane can have at most one common perpendicular.

**Problem 1.4.** Prove that a line and a boundary parallel to it cannot have a common perpendicular.

**Problem 1.5.** Prove that if (1) line $n$ passes through a point $P$ not on line $m$ and is boundary parallel to $m$, and (2) $Q$ is the foot of the perpendicular from $P$ to $m$, then $m$ is the boundary parallel to $n$ on the same side of $PQ$ on which $n$ is boundary parallel to $m$.

**Problem 1.6.** Derive the two kinds of geodesics on the pseudo-sphere from the element of arc length.

**Problem 1.7.** For a hyperbolic right triangle having an acute angle $\theta$ with adjacent side $X$ and hypotenuse $H$, show that

$$\cos\theta = \frac{\tanh(X/k)}{\tanh(H/k)}.$$

**Problem 1.8.** We can develop analytic geometry for the hyperbolic plane of curvature $-k^2$ by considering a pair of mutually perpendicular axes labeled the $X$-axis and the $Y$-axis, as shown in Fig. 16. From a point in the upper half-plane, we drop lines perpendicular to the two axes and use the distances from the feet of those perpendiculars to the origin as the coordinates of the point. The resulting quadrilateral has three right angles (and hence necessarily an acute angle at the point being labeled) and is usually called a *Lambert quadrilateral* after Johann Heinrich Lambert (1728–1777), although such quadrilaterals had been studied centuries earlier by Thabit ibn-Qurra (826–901). Show that the conversions from polar

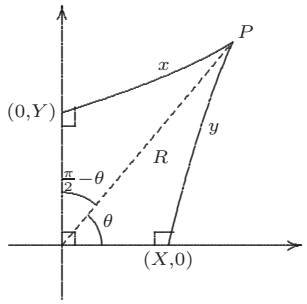


FIGURE 16. The basis of analytic geometry in the hyperbolic plane: A Thabit–Lambert quadrilateral. The point $P$ has rectangular coordinates $(X, Y)$ and polar coordinates $(R, \theta)$. The angle at $P$ is acute, and sides $x$ and $y$ are longer than the sides opposite them.

coordinates $(R, \theta)$ to rectangular coordinates $(X, Y)$ and vice versa are given by

$$\begin{aligned} X &= k \operatorname{arctanh}\big(\tanh(R/k)\cos\theta\big)\,, \\ Y &= k \operatorname{arctanh}\big(\tanh(R/k)\sin\theta\big)\,, \\ R &= k \operatorname{arctanh}\sqrt{\tanh^2(X/k)+\tanh^2(Y/k)}\,, \\ \theta &= \arctan\Big(\frac{\tanh(Y/k)}{\tanh(X/k)}\Big)\,. \end{aligned}$$

**Problem 1.9.** Show that the element of arc length in polar coordinates $(R, \theta)$ in the hyperbolic plane is

$$ds^2 = dR^2 + k^2 \sinh^2\Big(\frac{R}{k}\Big)\, d\theta^2\,.$$

**Problem 1.10.** Deduce from the previous problem that this means that the circumference of a hyperbolic circle of radius $R$ is $2\pi k \sinh\big(R/k\big)$, and that the element of area is

$$dA = k \sinh\Big(\frac{R}{k}\Big)\, dR\, d\theta\,,$$

so that the area of a circle of radius $R$ is $2\pi k^2\Big(\cosh\big(R/k\big) - 1\Big)$

**Problem 1.11.** Show that the geodesics in the disk of radius $c$ with the relativistic velocity metric of Notebook 9, that is,

$$ds^2 = \frac{c^4}{(c^2 - r^2)^2}\, dr^2 + \frac{c^2 r^2}{c^2 - r^2}\, d\theta^2,$$

are precisely the chords (not including their endpoints).

**Problem 1.12.** Verify that along the geodesic on the pseudo-hemisphere that projects into the equatorial plane as the curve whose polar equation is

$$r = \frac{k r_0}{\sqrt{k^2 - r_0^2(\theta - \theta_0)^2}}$$

the element of arc length is

$$ds = \frac{k^2 r_0}{k^2 - r_0(\theta - \theta_0)^2}\, d\theta\,,$$

where $\theta_0$ is the value of $\theta$ at which the radial coordinate $r$ has its minimum value ($r_0$) along the geodesic (which corresponds to the highest point on the geodesic itself).

## APPENDIX 2

# Calculus of Variations. Geodesics

Because the dynamics of general relativity requires the notion of geodesic paths, we assemble here the basic information about these paths, beginning with the very simple second-order differential equation for paths that minimize any particular function. To apply this differential equation in geometry, we need differential expressions for the local length of a path. For that purpose, we derive the first fundamental form for paths on a surface in $\mathbb{R}^3$ and the infinitesimal element of area on the surface.

### 1. Euler's Equation

In this section we give an informal derivation of the important differential equations used in the calculus of variations to find the path $\big(x(t), y(t)\big)$ that minimizes (or maximizes) an integral

$$I(x, y, x', y', t) = \int_{t_1}^{t_2} F\big(x(t), y(t), x'(t), y'(t), t\big)\, dt\,,$$

In most textbooks on the calculus of variations (not all), the derivatives here are denoted by dots above the symbol denoting the function that is being differentiated with respect to $t$. That usage was established by Lagrange in the eighteenth century, rescuing a notation due to Newton that the British themselves were on the verge of abandoning in favor of the Leibnizian "differential coefficient" $d/dt$.

The path is not determined by the Euler equation(s) alone. Generally, one needs to prescribe in addition either the values of the two ordered pairs $\big(x(t_1), y(t_1)\big)$ and $\big(x(t_2), y(t_2)\big)$ (a boundary-value problem) or the values of the two ordered pairs $\big(x(t_1), y(t_1)\big)$ and $\big(x'(t_1), y'(t_1)\big)$ (an initial-value problem) in order to determine the minimizing functions $x(t)$ and $y(t)$ completely.

**Remark 2.1.** We are actually going to find only a *necessary* condition for an extremum. Sufficient criteria do exist, and the reader can probably infer what they are, by merely imitating the standard argument that we are going to give below, only using second-order derivatives rather than first-order ones.

We are also not going to worry whether the extremum we find is a maximum or a minimum. The application of it is tested by experimental measurements of its predictions, and is not the concern of the present book. Textbooks on the calculus of variations are always careful to note that a great-circle arc on a sphere is the shortest path between its endpoints if it is less than 180° long. If it is longer than that, however, it is the *rest* of its great circle that forms the shortest path. Nevertheless, even great-circle arcs longer than 180° are the shortest paths joining their endpoints *in their immediate neighborhood*. That is to say, any "nearby" path will be at least as long. (The complementary arc of the great circle, which is the shortest path,

is not "nearby.") Here we are following standard usage in applications, where the underlying principle is that the integral is to have a "stationary" value (where Euler's equation is satisfied), not necessarily a minimum or maximum. Historically, however, it was usually assumed that the value is a minimum, and we shall find it convenient to talk as if it always is a minimum.

Without being too fussy about the domain of the function $F$, we shall assume that $x$ and $y$ have continuous derivatives on the closed interval $[t_1, t_2]$ (including one-sided derivatives at $t_1$ and $t_2$), and that they can be smoothly perturbed by adding to them "small" functions $u(t)$ and $v(t)$ satisfying $u(t_1) = u(t_2) = v(t_1) = v(t_2) = 0$ for the boundary-value problem or $u(t_1) = v(t_1) = u'(t_1) = v'(t_1) = 0$ for the initial-value problem and that the new point $\big(x(t)+ru(t), y(t)+sv(t), x'(t)+ru'(t), y'(t)+sv'(t), t\big)$ also belongs to the domain of $F$ for $-1 \le r \le 1$ and $-1 \le s \le 1$. If $\big(x(t), y(t)\big)$ minimizes the integral $I$, then for a small function $u(t)$ satisfying the extra conditions and for $-1 \le r \le 1$ we have

$$\int_{t_1}^{t_2} \big(F\big(x(t) + ru(t), y(t), x'(t) + ru'(t), y'(t), t\big) - F\big(x(t), y((t), x'(t), y'(t), t\big)\, dt \ge 0\,.$$

By Taylor's theorem on the variable $r$, this means

$$\int_{t_1}^{t_2} r\big(u(t)D_1F\big(x(t), y(t), x'(t), y'(t), t\big)+u'(t)D_2F\big(x(t), y(t), x'(t), y'(t), t)\big)+o(r)\, dt \ge 0\,,$$

where $o(r)$ denotes a quantity such that $o(r)/r \to 0$ as $r \to 0$. Dividing this inequality by a positive value of $r$ and then letting $r \downarrow 0$, we conclude that

$$\int_{t_1}^{t_2} u(t)D_1F\big(x(t), y(t), x'(t), y'(t), t\big) + u'(t)D_3F\big(x(t), y(t), x'(t), y'(t), t\big)\, dt \ge 0\,,$$

Integrating the second term by parts, using the assumption that $u(t_1) = 0 = u(t_2)$, we get

$$\int_{t_1}^{t_2} u(t)\big(D_1F\big(x(t), y(t), x'(t), y'(t), t\big) - \frac{d}{dt}D_3F(\big(x(t), y(t), x'(t), y'(t), t)\big)\, dt \ge 0$$

for all small perturbations $u(t)$. Replacing $u$ by $-u$, we see that in fact

$$\int_{t_1}^{t_2} u(t)\big(D_1F\big(x(t), y(t), x'(t), y'(t), t\big) - \frac{d}{dt}D_3F(\big(x(t), y(t), x'(t), y'(t), t)\big)\, dt = 0\,.$$

(The same reasoning would apply if we assumed that the integral was maximized rather than minimized.)

This equation must hold for all small functions $u(t)$ satisfying $u(t_1) = u(t_2) = 0$ or $u(t_1) = u'(t_1) = 0$ and having a continuous derivative $u'(t)$ on the interval $[t_1, t_2]$. Now it is not difficult to show that if $f(t)$ is any continuous function on an interval $[a, b]$ such that

$$\int_a^b u(t)f(t)\, dt = 0$$

for all small continuously differentiable functions $u(t)$ satisfying $u(a) = u(b) = 0$ or $u(a) = u'(a) = 0$, then $f(t) \equiv 0$ on that interval (Problem 2.1 below). Thus, we arrive at the Euler equation

$$D_1F\big(x(t), y(t), x'(t), y'(t), t\big) - \frac{d}{dt}D_3F\big(x(t), y(t), x'(t), y'(t), t\big) = 0\,,$$

which is stated in classical notation in the form

$$\frac{d}{dt}\Big(\frac{\partial F}{\partial x'}\Big) = \frac{\partial F}{\partial x}\,.$$

The analogous equation for $y$ is an immediate consequence when $x$ and $x'$ are replaced by $y$ and $y'$.

This result can be generalized to multiple integrals. Thus, suppose we wish to minimize the integral

$$I(z, z_x, z_y, x, y) = \iint\limits_U F(z, z_x, z_y, x, y)\,dx\,dy\,.$$

(Here $z_x$ and $z_y$ are the partial derivatives of $z$ with respect to $x$ and $y$.)

The same reasoning as above leads to Euler's equation, which in classical form is

$$\frac{\partial F}{\partial z} = \frac{\partial}{\partial x}\,\frac{\partial F}{\partial z_x} + \frac{\partial}{\partial y}\,\frac{\partial F}{\partial z_y}\,.$$

Thus, if we wish to minimize the area $A$ of the graph of a function $z(x, y)$ above a region $U$, we have the integral

$$A(z_x, z_y) = \iint\limits_U \sqrt{1 + z_x^2 + z_y^2}\,dx\,dy\,,$$

which leads to the equation

$$\frac{\partial}{\partial x}\bigg(\frac{\frac{\partial z}{\partial x}}{\sqrt{1 + z_x^2 + z_y^2}}\bigg) + \frac{\partial}{\partial y}\bigg(\frac{\frac{\partial z}{\partial y}}{\sqrt{1 + z_x^2 + z_y^2}}\bigg) = 0\,.$$

This equation was derived by Lagrange in 1762. It can be rewritten as

$$\nabla^2 z + z_y^2\frac{\partial^2 z}{\partial x^2} - 2z_x\,z_y\,\frac{\partial^2 z}{\partial x\,\partial y} + z_x^2\frac{\partial^2 z}{\partial y^2} = 0\,,$$

where $\nabla^2 z$ is the Laplacian (see Chapter 6 of Volume 1).

If $z$ also depends on time $t$ and the surface represents a vibrating membrane for which the amplitude of the vibrations is very small (that is, $z_x$ and $z_y$ are small), we get approximately the equation

$$\nabla^2 z = 0\,.$$

In general, the linearized equation of a vibrating membrane whose amplitude at time $t$ is $z(t; x, y)$ is

$$\frac{\partial^2 z}{\partial t^2} = c^2\nabla^2 z\,.$$

It would seem, then that a membrane in the shape of the graph of a *harmonic function*, for which $\nabla^2 z = 0$, should be stationary: It shouldn't vibrate at all.

## 2. Length and Area on a Surface

Consider a surface $S$ in $\mathbb{R}^3$. In ancient times, such an object was one of the most difficult to study, and in most cultures, only polyhedra and spheres were considered. The Greeks did a little better, considering ellipsoids and paraboloids of revolution. But the tools available to geometers—circles, lines, planes, polyhedra, cones, cylinders, spheres, and the like—were inadequate to study more than a very limited collection of surfaces (or, for that matter, curves). Not until Descartes and Fermat showed how algebra could be applied in geometry (what we now call analytic geometry) did it become possible to study any very general surfaces. While one can visualize many surfaces, rigorous proofs of their properties are nearly impossible to achieve using Euclidean methods. We are forced to resort to analytic representations, and that usually means parametrizations.

The idea of representing a curve as the image of a line segment under a continuously differentiable mapping or a surface as the image of a plane region, again under a continuously differentiable mapping, has proved to be very fruitful. It is the ultimate source of our modern concept of a manifold (see Appendix 4). Euler and Gauss achieved spectacular insights into some very complex curves and surfaces by using parametrizations. The human ability to use symbols gave a much larger scope to our ability to think geometrically. Along with the possibility of studying extremely complex curves and surfaces came the ability to think rationally about spaces of any number of dimensions. For all these reasons, parametrizations have been a great boon to geometers. There are, however, certain disadvantages in using them:

1. Symbols are not pictures, and it takes some skill to learn to associate a set of formulas with a geometric object.
2. The parametrization of a curve or surface is not unique, and it may not be obvious that a property proved using a parametrization is intrinsic to the geometric object that is parameterized and not a peculiarity of the particular parametric representation of it.
3. Because parametrization is not unique, the problem arises of deciding when two mappings of parameter spaces actually define the same geometric object.
4. Ideally, we need the mapping from the parameter space to be continuously differentiable and to have an inverse of the same type. But in all of the most important examples, there are places where the mapping becomes many-to-one or has a singular differential (the linear mapping that approximates it best is not invertible).

The last three of these problems are solved by passing from parametrizations to manifolds. A manifold consists of a bundle of parametrizations, each one with its domain restricted so that it is one-to-one and has a continuously differentiable inverse. Since the manifold contains all possible parametrizations (or all possible parametrizations whose transition mappings have positive Jacobian in the case of an oriented manifold), nothing about it depends on any particular one of the parametrizations. The domains of the parametrizations overlap in such a way that if a point of the surface is imagined to be moving toward the edge of the range of one parametrization, approaching a point that is eithere not in the range of the parametrization or is the image of a point in the parameter space where the mapping has a singularity, the moving point will, before arriving at the "bad" point,

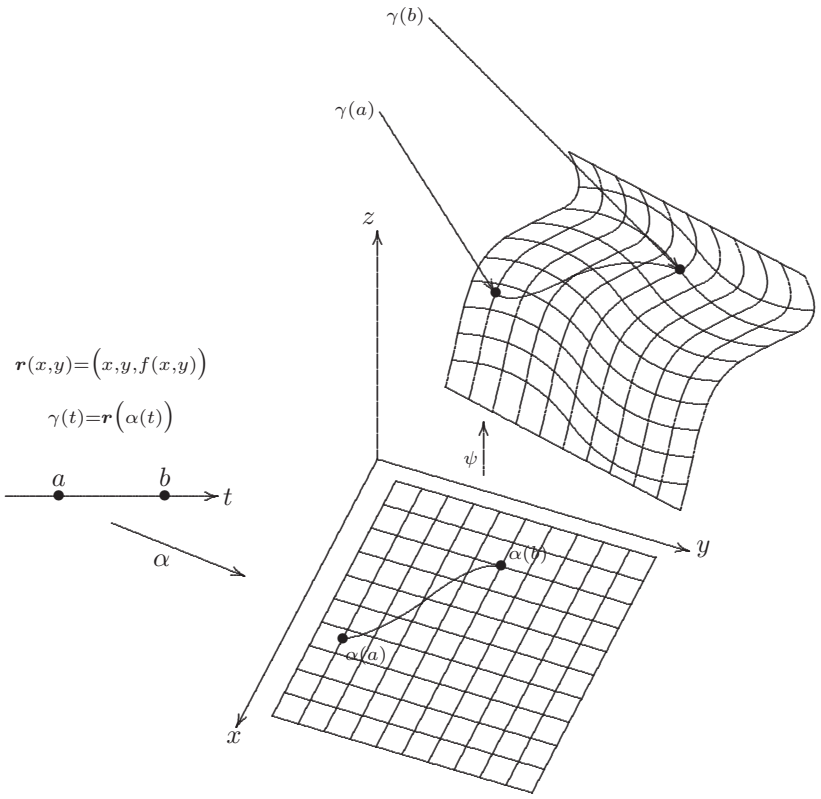


FIGURE 1. A parameterized curve $\gamma(t)$ on a surface $z = f(x, y)$ in $\mathbb{R}^3$. Since the points of the surface are parameterized by their projections into the $xy$-plane, the curve $\gamma(t)$ is the composition of the mapping $\boldsymbol{r}(x, y) = \big(x, y, f(x, y)\big)$ with a mapping $\alpha : [a, b] \to \mathbb{R}^2$. It is usually simpler to define the mapping $\alpha$ rather than to try to define $\gamma(t)$ directly as a point on a given surface in $\mathbb{R}^3$.

be inside the range of a second parametrization, which has an inverse defined at all points near the point in question.

Still, it is a nuisance to have to switch from one coordinate chart to another, and very often we prefer simply to use a "defective" chart that is singular or not one-to-one at a particular point or on a particular line. This informal approach can be justified by various theorems on termwise integration or differentiation of a sequence, but mathematicians often do not bother with such troublesome details. (Physicists and engineers almost never do.)

**2.1. Vector-valued parametrizations.** Using vector-valued parametrizations, we can study arc length, curvature, and area on some very complicated surfaces in $\mathbb{R}^3$, and the whole theory is applicable to $\mathbb{R}^n$ with very little complication. For now, we confine ourselves to the case of surfaces in $\mathbb{R}^3$. We represent such a surface as a mapping $\boldsymbol{r}$ from a region of the plane into $\mathbb{R}^3$, as shown in Fig. 1. The case shown there, in which the surface is the graph of a function $f(x, y)$ is sufficiently

general. (See the following section for justification of that remark.) The present discussion repeats some of what is discussed in Chapter 5, to which the reader is directed for more detailed explanations wherever necessary.

$$(x, y) \mapsto \boldsymbol{r}(x, y) = \big(x, y, f(x, y)\big)\,.$$

Since the theory we are about to develop applies to any vector-valued mapping $\boldsymbol{r}(u, v) = \big(x(u, v), y(u, v), z(u, v)\big)$, not only the case where $x(u, v) = u$ and $y(u, v) = v$, we shall not make any use of the assumption of this special form. By the Euclidean Pythagorean theorem, infinitesimal increments $dx$, $dy$, and $dz$ produce an infinitesimal change in arc length $ds$ given by $ds^2 = dx^2 + dy^2 + dz^2$. Since

$$\begin{aligned} dx &= \frac{\partial x}{\partial u}\,du + \frac{\partial x}{\partial v}\,dv \\ dy &= \frac{\partial y}{\partial u}\,du + \frac{\partial y}{\partial v}\,dv \\ dz &= \frac{\partial z}{\partial u}\,du + \frac{\partial z}{\partial v}\,dv\,, \end{aligned}$$

we can conveniently write this relation as

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2\,,$$

where

$$\begin{aligned} E &= \frac{\partial \boldsymbol{r}}{\partial u} \cdot \frac{\partial \boldsymbol{r}}{\partial u} = \Big|\frac{\partial \boldsymbol{r}}{\partial u}\Big|^2\,, \\ F &= \frac{\partial \boldsymbol{r}}{\partial u} \cdot \frac{\partial \boldsymbol{r}}{\partial v}\,, \\ G &= \frac{\partial \boldsymbol{r}}{\partial v} \cdot \frac{\partial \boldsymbol{r}}{\partial v} = \Big|\frac{\partial \boldsymbol{r}}{\partial v}\Big|^2\,. \end{aligned}$$

The quadratic form $ds^2$ was called by Gauss the *first fundamental form* of the surface. As is obvious from its definition, it can be used to calculate the arc length of any parameterized curve that lies on the surface, such as the curve $\gamma(t)$ shown in Fig. 1. All that is needed is to express $u$ and $v$ as functions of the curve parameter $t$, as is done using the mapping $\alpha(t)$ in the figure. Then $du = u'(t)\,dt$, $dv = v'(t)\,dt$, and so we have

$$ds = \sqrt{E\,\big(u'(t)\big)^2 + 2F\,u'(t)v'(t) + G\,\big(v'(t)\big)^2}\,dt\,.$$

**Remark 2.2.** The form $ds^2$ appears in generalized form in the differential geometry of manifolds. In Chapter 4, we introduced the metric coefficients $g_{ij}$ in four-dimensional space. In the two-dimensional parameter space of $u$ and $v$, the coefficients $E$, $F$, and $G$ give the metric in exactly the same way: $g_{11} = E$, $g_{12} = g_{21} = F$, and $g_{22} = G$. In contrast to the metric of space-time, the matrix of the classical metric, that is,

$$M = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}\,,$$

has the property of being *positive-definite*, that is, (1) it is symmetric and (2) for any non-zero vector $(s, t)$ the product

$$\begin{pmatrix} s & t \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix}$$

is positive.

Geometrically this means that the angle between $(s,t)$ and $(s',t')$, where

$$\begin{pmatrix} s' \\ t' \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix},$$

is acute.

A symmetric matrix $M$ for which this angle is always obtuse is *negative-definite*. Obviously, most matrices are not sign-definite. A necessary and sufficient condition for a $2 \times 2$ matrix to be positive-definite is that its determinant be positive and that its upper left (or lower right) entry be positive. (When the determinant is positive, these two entries have the same sign.) The positivity of this matrix is equivalent to the Schwarz inequality, which says that $(\boldsymbol{u} \cdot \boldsymbol{v})^2 < (\boldsymbol{u} \cdot \boldsymbol{u})(\boldsymbol{v} \cdot \boldsymbol{v})$ for any two non-collinear vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. For the matrix to be negative-definite, this determinant must be positive and the upper left or lower right entry must be negative. Obviously, that cannot happen when the upper left entry is $\boldsymbol{u} \cdot \boldsymbol{u}$ and the lower right is $\boldsymbol{v} \cdot \boldsymbol{v}$. A positive-definite metric is called *Riemannian*, after Bernhard Riemann (1826–1866).

The partial derivatives $\partial \boldsymbol{r}/\partial u$ and $\partial \boldsymbol{r}/\partial v$ represent vectors whose directions are tangent to the surface and whose magnitudes are respectively the factors by which an infinitesimal lengths $du$ and $dv$ along horizontal and vertical lines in the parameter space are multiplied under the mapping $(u,v) \mapsto \boldsymbol{r}(u,v)$. Since the infinitesimal element of area in the parameter space is $dA = du\,dv$, we should expect that the area of the parallelogram spanned by the vectors $\partial \boldsymbol{r}/\partial u\,du$ and $\partial \boldsymbol{r}/\partial v\,dv$ would represent the infinitesimal element of surface area $dS$.

That area is just the magnitude of the cross product of the two vectors. Since $du\,dv$ can be factored out of this cross product, the result is that the magnitude of the cross product of the partial derivatives is the factor by which the infinitesimal area $du\,dv$ is multiplied under the mapping. Using the well-known equation

$$|\boldsymbol{u} \times \boldsymbol{v}|^2 + (\boldsymbol{u} \cdot \boldsymbol{v})^2 = |\boldsymbol{u}|^2 |\boldsymbol{v}|^2,$$

we find that the element of surface area is

$$dS = \left|\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}\right| du\,dv = \sqrt{\left|\frac{\partial \boldsymbol{r}}{\partial u}\right|^2 \left|\frac{\partial \boldsymbol{r}}{\partial v}\right|^2 - \left(\frac{\partial \boldsymbol{r}}{\partial u} \cdot \frac{\partial \boldsymbol{r}}{\partial v}\right)^2}\, du\,dv = \sqrt{EG - F^2}\, du\,dv\,.$$

## 3. Euler's Geodesic Law

*It is readily apparent that only motion free of perturbation from the resistance of the medium is being considered here, and the reason is easy to see: When there is resistance, a body arriving at the same point over different paths does not acquire the same speed. For this reason, apart from any resistance in the motion of the bodies being moved, it will always turn out that the sum of all the infinitesimal movements will be minimal. Indeed, this property holds not only in the motion of a single body, but also in the joint motion of more than one. No matter how they interact with one another, the total amount of movement of the bodies is always minimal. Since this motion is difficult to calculate, it is more easily understood from basic principles than from the results of calculation using this method. All bodies, due to inertia, resist any change of state. When subjected to a force, constrained bodies submit to a lesser degree than they would if they were unconstrained. As a result, the motion produced by the forces is necessarily less than when the body or bodies interact in any other way. This reasoning is not quite conclusive. Nevertheless, I have no doubt at all that it will surely lead to more evidence for, and a clearer formulation of, the principles of a sounder metaphysics, a project that I leave to those who have use for metaphysics.*

Leonhard Euler, *Methodus inveniendi lineas curvas*, p. 320. My translation.

Euler gave the first detailed and articulated formulation of a least-action principle that turned out to be not only an elegant way of formulating the (basically Aristotelian) mechanics of Kepler, Galileo, Descartes, and Newton but also a fruitful source of new hypotheses. It follows from his method that a particle moving over a surface in $\mathbb{R}^3$ and subject to no forces tangential to that surface—that is, the tangential component of force acting on it is zero—will move along a geodesic on the surface at a constant linear speed. This result appears to fit into the progression of ideas from Fermat's minimal principle for the propagation of light through Lagrange's formulation of Newton's laws in the language of the calculus of variations right down to Einstein's statement of his law of gravity. Accordingly, we shall devote the present section to an informal derivation of it.

The constancy of the linear speed, whatever path a particle is following, is a consequence of Newton's second law of motion, interpreted as a vector law: If a component of the force is zero, then the corresponding component of the velocity is constant. If the force is always perpendicular to the velocity, then the speed (the magnitude of the velocity vector) is constant, and conversely. That is, for motion having a continuous acceleration, the equation $\boldsymbol{v} \cdot \boldsymbol{v} = k$ for some constant $k$ is equivalent to the equation $\boldsymbol{v} \cdot \boldsymbol{v}' = 0$, which says that the velocity and the acceleration (force) are mutually perpendicular. Although a force perpendicular to the velocity is not necessarily perpendicular to a surface on which the body is moving, we have at least a sufficient condition: *If* the force on a particle is always perpendicular to a surface on which a particle is moving, then the linear speed of that particle will be constant.

Consider now a surface whose equation is $G(x, y, z) = 0$. A particle at point $\boldsymbol{r}(t) = \big(x(t), y(t), z(t)\big)$ may happen to be on that surface at an instant $t = t_0$, but that fact by itself does not imply anything significant, since the point may be moving transversally to the surface and just "accidentally" crossing it at that instant. We would say it is moving *along* the surface if $G\big(\boldsymbol{r}(t_0)\big) = 0$ and $\boldsymbol{r}'(t_0)$ is tangential to

the surface, which is to say, perpendicular to the gradient: $\nabla G\big(\boldsymbol{r}(t_0)\big) \cdot \boldsymbol{r}'(t_0) = 0$. The question we are interested in is the following: Under what conditions will the particle remain on the surface over an interval of time containing $t_0$? Obviously, if it remains on the surface, then $G\big(\boldsymbol{r}(t)\big) \equiv 0$, and differentiating with respect to $t$, we find that $\nabla G\big(\boldsymbol{r}(t)\big) \cdot \boldsymbol{r}'(t) \equiv 0$. This is a necessary condition for the particle to remain on the surface. Conversely, if $G\big(\boldsymbol{r}(t_0)\big) = 0$ and $\nabla G\big(\boldsymbol{r}(t)\big) \cdot \boldsymbol{r}'(t) \equiv 0$, then $G\big(\boldsymbol{r}(t)\big)$ has the constant value 0, and so these last two equations, taken together, are a sufficient condition for the particle to remain on the surface.

We now consider the case of a particle at position $\boldsymbol{r}(t)$ at time $t$ that is moving along the surface at some instant $t_0$. That is, $G\big(\boldsymbol{r}(t_0)\big) = 0$ and $\nabla G\big(\boldsymbol{r}(t_0)\big) \cdot \boldsymbol{r}'(t_0) = 0$. A particle can move quite chaotically over the surface if the forces acting on it are suitably chosen. It may speed up or slow down. But we can see easily that it will move at a constant rate of speed—that is, $\boldsymbol{r}'(t) \cdot \boldsymbol{r}'(t)$ will be constant—if there are no tangential forces acting on it. To formulate that condition, we assume that there is a scalar-valued function of position $\varphi\big(\boldsymbol{r}\big)$ such that

$$\boldsymbol{r}''(t) = \varphi\big(\boldsymbol{r}(t)\big) \nabla G\big(\boldsymbol{r}(t)\big) \tag{1}$$

at all times $t$

Taken together with the condition for the particle to remain on the surface ($\nabla G\big(\boldsymbol{r}(t)\big) \boldsymbol{r}'(t) \equiv 0$), this condition implies that $\boldsymbol{r}''(t) \cdot \boldsymbol{r}'(t) \equiv 0$, and hence that $\boldsymbol{r}'(t) \cdot \boldsymbol{r}'(t)$ is constant, as asserted. We now make one more bold claim about such a motion, stating a theorem proved by Euler in 1744:

**Theorem 2.1.** *Let $\boldsymbol{r}(t)$ be a vector-valued function that is twice continuously differentiable on an interval $(a, b)$ containing $[t_0, t_1]$ and such that (1) $G\big(\boldsymbol{r}(t_0)\big) = 0$ and (2) $\nabla G\big(\boldsymbol{r}(t)\big) \cdot \boldsymbol{r}'(t) \equiv 0$ for $t_0 \le t \le t_1$. If in addition the scalar-valued function $\varphi$ is such that Eq. (1) holds, then the path $\boldsymbol{r}(t)$, $t_0 \le t \le t_1$ is a geodesic on the surface $G(x, y, z) \equiv 0$, and the parameter $t$ is directly proportional to arc length along that path.*

Proof. We find easily that

$$0 = \frac{d}{dt}\Big(\nabla G\big(\boldsymbol{r}(t)\big) \cdot \boldsymbol{r}'(t)\Big) = \nabla G\big(\boldsymbol{r}(t)\big) \cdot \boldsymbol{r}''(t) + \boldsymbol{r}'(t) \cdot H\big(\boldsymbol{r}'(t)\big)\,,$$

where $H$ is the Hessian operator whose matrix is the tableau of second-order derivatives:

$$H\big(x, y, z\big) = \Big(\frac{\partial^2 G}{\partial x^2}x + \frac{\partial^2 G}{\partial x \partial y}y + \frac{\partial^2 G}{\partial x \partial z}z\,, \\ \frac{\partial^2 G}{\partial y \partial x}x + \frac{\partial^2 G}{\partial y^2}y + \frac{\partial^2 G}{\partial y \partial z}z\,,\ \frac{\partial^2 G}{\partial z \partial x}x + \frac{\partial^2 G}{\partial z \partial y}y + \frac{\partial^2 G}{\partial z^2}z\Big)\,.$$

Now by Eq. (1)

$$\nabla G\big(\boldsymbol{r}(t)\big) \cdot \boldsymbol{r}''(t) = \varphi\big(\boldsymbol{r}(t)\big)\big|\nabla G\big(\boldsymbol{r}(t)\big)\big|^2\,.$$

Therefore, the function $\varphi\big(\boldsymbol{r}(t)\big)$ has to be such that

$$\varphi\big(\boldsymbol{r}(t)\big) = -\frac{\boldsymbol{r}'(t) \cdot H\big(\boldsymbol{r}'(t)\big)}{\big|\nabla G\big(\boldsymbol{r}(t)\big)\big|^2}\,.$$

In order to see that this condition implies the Euler equations for a geodesic, we need to get the metric on the surface $G(x, y, z) \equiv 0$. The easiest way to do this is to note that on this surface

$$0 = dG(x, y, z) = \frac{\partial G}{\partial x}\, dx + \frac{\partial G}{\partial y}\, dy + \frac{\partial G}{\partial z}\, dz\,.$$

Since we are assuming the surface has no singularities, at least one of the three partial derivatives in this expression is non-zero. Without loss of generality, let us assume that it is $\partial G/\partial z$. We thus have

$$dz = -\frac{\frac{\partial G}{\partial x}}{\frac{\partial G}{\partial z}}\, dx - \frac{\frac{\partial G}{\partial y}}{\frac{\partial G}{\partial z}}\, dy\,,$$

so that

$$ds^2 = dx^2 + dy^2 + dz^2 = \left(1 + \left(\frac{\frac{\partial G}{\partial x}}{\frac{\partial G}{\partial z}}\right)^2\right) dx^2 + 2\frac{\frac{\partial G}{\partial x}}{\frac{\partial G}{\partial z}}\,\frac{\frac{\partial G}{\partial y}}{\frac{\partial G}{\partial z}}\, dx\, dy + \left(1 + \left(\frac{\frac{\partial G}{\partial y}}{\frac{\partial G}{\partial z}}\right)^2\right) dy^2\,.$$

This expression is messy. But, under our assumptions, the equation $G(x, y, z) \equiv 0$ is *locally* equivalent to the equation $f(x, y) - z \equiv 0$ for some continuously differentiable function $f(x, y)$. We now assume that $G(x, y) = f(x, y) - z$, so that $\partial G/\partial z = -1$, $\partial G/\partial x = \partial f/\partial x$, and $\partial G/\partial y = \partial f/\partial y$.

We then have

$$\begin{aligned} \nabla G(x, y, z) &= \frac{\partial f}{\partial x}\boldsymbol{i} + \frac{\partial f}{\partial y}\boldsymbol{j} - \boldsymbol{k}\,, \\ \boldsymbol{r}'(t) \cdot H\boldsymbol{r}'(t) &= \frac{\partial^2 f}{\partial x^2}\big(x'(t)\big)^2 + 2\frac{\partial^2 f}{\partial x \partial y} x'(t) y'(t) + \frac{\partial^2 f}{\partial y^2}\big(y'(t)\big)^2\,. \end{aligned}$$

. To save some writing, we shall use the abbreviation $\boldsymbol{r}'(t) \cdot H$
$big(vr'(t)\big) = \Delta$ in the argument below.

We intend to show that when the function $\varphi\big(\boldsymbol{r}(t)\big)$ satisfies the equation given above, the particle moves along a geodesic on the surface. (We already know that it stays on the surface and moves at a constant linear speed.) Because the speed is constant, we can use the infinitesimal element of arc length $ds$ interchangeably with the infinitesimal element of time $dt$.

Since $dz = \frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy$, we have

$$ds^2 = dt^2 = \left(1 + \left(\frac{\partial f}{\partial x}\right)^2\right) dx^2 + 2\frac{\partial f}{\partial x}\,\frac{\partial f}{\partial y}\, dx\, dy + \left(1 + \left(\frac{\partial f}{\partial y}\right)^2\right) dy^2\,.$$

When we take arc length along a path (or time, in this case) as the parameter, the integrand in the integral that gives the arc length of the path is equal to 1 at every point of the path. That integrand is

$$G(x, y, x', y') = \sqrt{\left(1 + \left(\frac{\partial f}{\partial x}\right)^2\right)(x')^2 + 2\frac{\partial f}{\partial x}\,\frac{\partial f}{\partial y}\, x'\, y' + \left(1 + \left(\frac{\partial f}{\partial y}\right)^2\right)(y')^2}\,.$$

The relation $G\big(x(s), y(s), x'(s), y'(s)\big) \equiv 1$ greatly simplifies the computation of the two partial derivatives in each Euler equation, and we find that three terms cancel on each side of the equation. Without going into the details, which are a

routine computation, we find that

$$\begin{aligned}\left(1+\left(\frac{\partial f}{\partial x}\right)^2\right)x'' + \frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\,y'' &= -\frac{\partial f}{\partial x}\Delta\,,\\ \frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\,x'' + \left(1+\left(\frac{\partial f}{\partial y}\right)^2\right)y'' &= -\frac{\partial f}{\partial y}\Delta\,,\end{aligned}$$

where

$$\Delta = \frac{\partial^2 f}{\partial x^2}(x')^2 + 2\frac{\partial^2 f}{\partial x \partial y}x'\,y' + \frac{\partial^2 f}{\partial y^2}(y')^2\,.$$

It is then a trivial matter to solve the system of Euler equations for the first two components of the acceleration. We find that

$$\begin{aligned}x'' &= \frac{-\frac{\partial f}{\partial x}\Delta}{1+\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2}\,,\\ y'' &= \frac{-\frac{\partial f}{\partial y}\Delta}{1+\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2}\,,\end{aligned}$$

from which we conclude that

$$\frac{\partial f}{\partial y}\,x'' - \frac{\partial f}{\partial x}\,y'' = 0\,.$$

As for the third component of the acceleration $z''$, since $z = f(x,y)$, we easily compute that

$$z'' = \Delta + \frac{\partial f}{\partial x}\,x'' + \frac{\partial f}{\partial y}\,y''\,.$$

But, in view of our original system of Euler equations, this equation immediately implies that

$$\begin{aligned}\frac{\partial f}{\partial x}z'' &= -x''\,,\\ \frac{\partial f}{\partial y}z'' &= -y''\,.\end{aligned}$$

Putting these three equations together, we conclude that the cross product of the acceleration with the normal (gradient) vector is zero:

$$\nabla F(x,y,z)\times\left(x''\,,\,y''\,,\,z''\right) = \mathbf{0}\,.$$

In other words, the acceleration (and hence also the force) is parallel to $\nabla F$, and therefore perpendicular to the surface. Thus, the acceleration is indeed a scalar multiple of the gradient at each point, and the explicit computation shows that for a geodesic, the scalar coefficient is

$$\varphi(\boldsymbol{r}(t)) = \frac{-\Delta}{|\nabla G|^2} = \frac{-\boldsymbol{r}'(t)\cdot H\boldsymbol{r}'(t)}{|\nabla G|^2}\,,$$

which is exactly what we computed that it had to be in order to confine the particle to the surface. Thus, the condition for the particle to stay on the surface is the same as the condition for it to move along a geodesic of the surface, and Euler's theorem is now proved. □

**Remark 2.3.** This theorem is highly significant because of its extensions. It is stated for a particle moving over a surface in $\mathbb{R}^3$. The surface, however, need not be physical. It can be a purely geometric construct, and in that context, tangential forces can be regarded as "internal" to the surface and normal forces as purely "external." For a particle moving in three-dimensional space, the surface may arise as the integral of a differential equation, that is, a function $G(x, y, z)$ such that $G\big(x(t), y(t), z(t)\big)$ is constant throughout an interval of time. In that case, the surface is essentially constructed from the forces known to act on the particle. This theorem asserts that if the forces do nothing except confine the particle to a surface, that is, they do not accelerate the particle on the surface, then the particle "naturally" moves along a geodesic.

We were able to prove this theorem by making use of properties of $\mathbb{R}^3$, but $\mathbb{R}^3$ is an ambient space inside which the surface on which the particle "lives" is embedded. In the abstract, we might wish to consider a particle moving on a surface defined intrinsically and not embedded in any larger space. In that case, the notion of external forces makes no sense at all, and the absence of tangential forces becomes simply an absence of force altogether, and so the theorem would assert that a particle moving in the absence of forces always moves along a geodesic. That is the basic principle behind the Principle of Least Action, and it is the essential idea of Einstein's law of gravity.

There is yet more to be said, however, since the very idea of force presumes a Newtonian combination of geometry and physics that is not at all essential. Poincaré pointed out that the mixture of geometric and physical laws is not uniquely determined. We can use a variety of forms of geometry, provided we are willing to change our physical laws. That point is well illustrated by relativistic velocity triangles, which are derived from principles that belong to physics, but whose properties are those of triangles in hyperbolic geometry.

**3.1. Lagrangian mechanics.** Although Lagrangian mechanics is discussed in Chapters 2, 4, and 7 of Volume 1, we are inserting at this point a more systematic exposition in slightly different language in order to emphasize its connection with the calculus of variations. In the quotation at the beginning of this section, Euler gave the very good reason why he was not taking friction or fluid resistance into account: When these are present, the velocity a particle will acquire at a given point as a result of the action of a force depends not only on the force, but also on the path by which the body arrived at that point. Forces having the property that the velocity of the particle at a given point depends only on the velocity at the initial point of the path over which it traveled to the given point are said to be *conservative*. Many important forces in classical physics, such as those due to what we now call gravitational, electrical, and magnetic fields, are conservative. For these forces, Euler's analysis is the key to a very elegant and useful reformulation of Newton's laws, introducing ideas that were not explicit in the work of Newton himself. A brief historical digression may help to make this point clearer.

The intuitive basis of Newton's formulation of mechanics is not far removed from that of the ancient Greeks. In his *Physics*—the word in Greek means the *intrinsic nature of things*, that is, what they are "in and of themselves"—Aristotle laid down the rule that force (*dynamis*) is required to produce motion. Anything that moves must be moved by something else until, avoiding an infinite regression,

one arrives eventually at an eternally unmoved mover.[1] In Book VII of the *Physics*, he laid down the rule that the force moving a body is jointly proportional to its size and the distance moved in a fixed time. In our terms, if we interpret *size* to mean mass, that would make force proportional to what we now call momentum, that is, mass times velocity. Even though this explanation of motion does not begin to achieve as much as the Newtonian scheme, as a first attempt to mathematize the science of mechanics[2] it is quite admirable. From these principles, it follows that continuous force is required to sustain motion in relation to an absolute space. Given that the frictional and fluid resistance to motion occurs universally, one can perfectly well understand that it did not occur to Aristotle to abstract from them and imagine what the world would be like without them. There really is no perpetual motion in nature. Aristotle had a practical bent and tried to deal with the world as it actually was, just as engineers must necessarily do at all times when they are engaged in designing devices for human use.

It was Descartes who first clearly formulated the principle that motion is eternal in his 1644 work *The Principles of Philosophy*, saying

> *. . . it seems evident to me that it was none other than God who, in his omnipotence, created matter with the motion and rest of its parts and who now conserves by his regular operations in the universe exactly the same amount of movement and rest as was set out when he created it. For, though movement be only a form in inert matter, that matter nevertheless has a definite quantity of it, which never increases or decreases, even though there may be more or less of it in different parts. This is why, when one piece of matter moves twice as fast as another and the other is twice as large as the first, we must consider that there is just as much motion in the smaller as in the larger, and whenever the motion of one piece decreases, that of some other piece increases in proportion. . . if a body has once begun to move, we must conclude that it continues to move, and that it will never stop of its own accord. . . each piece of matter tends to continue its motion in straight lines, never in curves. . . .* [My translation from the French edition.]

Here you see Newton's first law of motion clearly stated in a work written while Newton was still a baby.[3] Moreover, the important principle of conservation

[1] Or a number of unmoved movers. Aristotle wasn't sure whether there were 47 of them or 55 of them. He did not invest the unmoved mover(s) with personality or any divine qualities, but naturally his disciples in the Christian Middle Ages found this concept very compatible with their theology.

[2] The Greek word *Mechanike* refers to *machinery*. It was the study of devices that produce effects *contrary to the intrinsic nature of things*. Machines make it possible for a small force to move a great weight, for example. Thus, in a sense physics and mechanics were the study of two opposite sets of phenomena. Copernicus, although he broke with Aristotle's cosmology, still accepted the Aristotelian distinction between natural motion and artificial motion. (See the quotation at the beginning of Section 12 of Chapter 1 in Volume 1.)

[3] I cannot help remarking at this point that Newton was contemptuous of Descartes' brilliant idea of algebraizing geometry. One does wonder if the title of his great work *Mathematical Principles of Natural Philosophy* was chosen to challenge Descartes' *Principles of Philosophy*. In trying to avoid the difficulties he had seen Galileo fall into by endorsing the Copernican cosmology, Descartes tied himself in knots with a theory of vortices, allowing that the earth surely does not move within its vortex, but the vortex does move. The notion of vortices—arising, as I believe, out of a desperate attempt to reconcile Ptolemaic astronomy with the now-obvious results of

of momentum is laid out, along with the principle that unforced motion is always along a straight line—that is, a geodesic in the Euclidean space that was the only one known at the time. Descartes still believed in a natural state of the universe, although he disagreed with Aristotle as to what that natural state was. His principle looks like only a slight modification of what Aristotle said, but the consequences of that modification—making force responsible for *changing*, rather than *sustaining*, momentum—were enormous.

In the light of what has just been said, we can say, broadly speaking, that the main alteration in mechanical principles between Aristotle and Newton is a matter of abstraction, that is, neglecting small effects that complicate matters so as to get a workable mathematical model. This kind of abstraction still goes on, as anyone who has ever taken a course of elementary physics will remember: In fact, pulleys are not weightless and frictionless, ropes are not infinitely thin, tensions and densities are not constant at every point of a vibrating string at all times, the period of a pendulum does depend on the amplitude of the arc it describes, and so on. These complications can be dealt with one at a time, but in order to get started, we need a mathematical model that works approximately, and that means one that neglects some aspects of a motion (temporarily). Natural philosophers have long been inclined to begin an analysis by using the simplest possible mathematical model: the linear model, as, for example, in Hooke's law of elasticity. To ancient natural philosophers like Aristotle and Archimedes, linearity meant stating physical principles in terms of direct proportion. Two variable quantities $x$ and $y$ are in direct proportion if their ratio remains constant. As we now phrase the matter algebraically, $y = kx$ for some constant $k$. This route was undoubtedly necessary for physics to get started, but of course many relations are not linear, most prominently the Pythagorean theorem. An early example of a nonlinear relation in physics is the law of refraction of light, for which Ptolemy gave what we now recognize as a quadratic relation between the angle of incidence and the angle of refraction. (The actual relation here is not even algebraic; it is trigonometric.) Nonlinear relations are difficult to handle without algebra, but the calculus greatly expanded the range of applicability of linear methods by the introduction of infinitesimal reasoning. On the infinitesimal level, all relations become linear, just as an infinitesimal portion of a curve is a straight line. Even with calculus, however, we still occasionally "linearize" some expressions such as the equation of pendulum motion to make them tractable, as we shall do below.

This digression now completed, let us return to what Euler was doing. When the decision is made to consider only conservative systems, those in which, as we now say, no work is done in moving a particle around a closed loop, the Newtonian analysis can be reformulated. The two sides of the equation $F = ma$ that expresses Newton's second law of motion can be thought of in an Aristotelian way as a fair exchange between a stimulus (the force $F$) imposed and response to it by a particle of mass $m$, which is the acceleration $a$. When velocity-dependent effects such as fluid resistance and mass-dependent effects such as friction are neglected, the force usually depends only on the location $\big(x(t), y(t), z(t)\big)$ of the particle at

telescopic observation—is truly muddled, and Newton demolished it quite effectively, although a lesser mind than his would have sufficed to do so. I suspect Descartes himself knew better but didn't dare say so. In one passage of his *Principles* he expresses at great length how willing he is to renounce the entire theory, leading the reader to wonder why he considered it necessary to be so emphatic about that point.

time $t$. Independently of the specific particle, one can arbitrarily choose a given point $(x_0, y_0, z_0)$ and consider the work required to move the particle from that point at time $t = 0$ to the general point $\big(x(t), y(t), z(t)\big)$. That work depends only on the location of the point and is independent of the path, It is expressed as a line integral in the language of differentials used by Euler:

$$W(x, y, z) = \int F_1\, dx + F_2\, dy + F_3\, dz\,,$$

where the force $F$ has components $(F_1, F_2, F_3)$. By the fundamental theorem of calculus, we have

$$\begin{aligned} F_1 &= \frac{\partial W}{\partial x}\,, \\ F_2 &= \frac{\partial W}{\partial y}\,, \\ F_3 &= \frac{\partial W}{\partial z}\,. \end{aligned}$$

For the sake of making a pedagogical point, we are now going to look at these equations anachronistically, in terms of a later language developed by William Rowan Hamilton (1805–1865) as quaternions and refined by British and American mathematicians into the subject of vectors. In that language, $\boldsymbol{F} = \nabla W$ and the integral is the integral of the differential $\boldsymbol{F} \cdot d\boldsymbol{r} = \boldsymbol{F}\big(\boldsymbol{r}(t)\big) \cdot \boldsymbol{r}'(t)\, dt$. It represents the sum of the infinitesimal amounts of work $dw$ done by the tangential component of the force $(\boldsymbol{F} \cdot \boldsymbol{r}'/|\boldsymbol{r}'|$ moving the particle an infinitesimal distance $ds = |\boldsymbol{r}'|\, dt$. As a consequence, when we parameterize the integral, we find that we are integrating the derivative of the composite function $W\big(x(t), y(t), z(t)\big)$, and that the result is the difference in the values of $W$ at the two end-points. If we ignore some potential complications due to the topology of the space, it follows that a conservative force field is the gradient of the scalar-valued work function $W(x, y, z)$. Since the point $(x_0, y_0, z_0)$ where $W$ vanishes is arbitrary, it is customary, to add a constant to $W$ so as to make the minimum amount of work equal to 0. (Adding a constant to $W$ will not effect the relation $\boldsymbol{F} = \nabla W$.) In that case, $W(x, y, z)$ is the maximum amount of work that the force can do in moving the particle to the point $(x, y, z)$, the maximum being taken over all possible starting points. For technical reasons, we prefer to work with the negative of the work, that is, the function $E_p(x, y, z) = -W(x, y, z)$ representing the minimum amount of work done against the force in moving the particle to the point $(x, y, z)$. The function $E_p(x, y, z)$ is called the *potential* of the force, that is, it represents the maximum amount of work that can be achieved using the force. Remembering that the gradient of a function points in the direction of maximum increase of the function, we can see that a force $\boldsymbol{F} = -\nabla E_p$ has the effect of "stimulating" a particle to decrease its potential by moving in the direction of most rapid decrease. That is no surprise to anyone: Water flows downhill and always flows in the direction of steepest descent. For the gravitational force, where

$$\boldsymbol{F}(\boldsymbol{r}) = -\frac{GMm}{r^3}\boldsymbol{r}\,,$$

$(r = \sqrt{\boldsymbol{r} \cdot \boldsymbol{r}} = \sqrt{x^2 + y^2 + z^2})$ we have

$$E_p(\boldsymbol{r}) = -\frac{GMm}{r} = -\frac{GMm}{\sqrt{x^2 + y^2 + z^2}}\,.$$

Thus, the maximum potential here is 0, to be achieved by moving the particle of mass $m$ to infinity.

But, taking into account the new principle that force represents the rate of change of momentum, we have $\boldsymbol{F}\big(\boldsymbol{r}(t)\big) = m\boldsymbol{r}''(t)$, and thus the infinitesimal amount of work done by the tangential component is $m\boldsymbol{r}''(t) \cdot \boldsymbol{r}'(t)\, dt$. When this differential is integrated, the result is a quite different scalar function, one that depends on the velocity rather than the position. Specifically, it is

$$E_k(\boldsymbol{r}) = \frac{1}{2} m\boldsymbol{r}' \cdot \boldsymbol{r}' \,.$$

The function $E_k$ is called the *kinetic energy*.[4]

With these definitions, we can reformulate Newton's laws for the motion of a particle from time $t_0$ to time $t_1$: The position $\big(x(t), y(t), z(t)\big)$ of the particle at time $t$ is such that the integral

$$\int_{t_0}^{t_1} E_k\big(x'(t), y'(t), z'(t)\big) - E_p\big(x(t), y(t), z(t)\big)\, dt$$

assumes a minimum value over all the possible paths.[5]

This integral is called the *action* integral, and the integrand is called the *Lagrangian*, after J.-L. Lagrange (1736–1813).[6] The reformulation of Newton's laws in terms of the Lagrangian is called the *principle of least action*. For the case of the motion of a particle under the gravitational attraction of the sun, Euler's equation says, for example, that

$$\frac{d}{dt}\frac{\partial E_k}{\partial x'} = -\frac{\partial E_p}{\partial x} \,,$$

which translates to

$$mx'' = -\frac{GMm}{\big(x^2 + y^2 + z^2\big)^{3/2}} x \,,$$

and with the analogous equations for $y$ and $z$ taken into account, gets us

$$\boldsymbol{F}\big(\boldsymbol{r}(t)\big) = m\boldsymbol{r}''(t) = -\frac{GMm}{r(t)^3}\boldsymbol{r}(t) \,,$$

which is exactly Newton's law of gravity.

In this way, we are led to formulate the law of gravity not as a stimulus-and-response kind of equation but as a principle of minimal action, in which kinetic energy and potential energy are exchanged by the minimum amount possible. True, we had to define the potential energy in terms of the Aristotelian notion of force; but once that is done, we need not inquire about the source of the potential function.

---

[4] The word *energy*, so familiar to us nowadays, was not common in the eighteenth century. It was rather an invented word, based on the Greek root *-erg*, meaning *work*. Etymologically, *energy* is the work "contained in" an object, and the two sides of Newton's second law give the two forms in which it exists: As potential energy depending on position, representing the work that can be extracted by allowing the object to move to a point of lower potential, like the falling water that turns a turbine, and the kinetic energy of motion, representing the work that can be extracted by causing the body to stop moving, like the expanding molecules of gas in an internal combustion engine.

[5] Actually, we won't have such an all-encompassing result as this statement implies. The path will be a *local* extreme value for the time integral of the energy, but not necessarily a minimum and not necessarily a *global* extreme value. In any case, physicists now state this principle in terms of a *stationary value* of the integral. As stated above, we are going to ignore these complications.

[6] Known from his publications at the Paris Academy of Sciences as Joseph-Louis Lagrange. He was actually Italian, and his birth name was Giuseppe Lodovico Lagrange.

As far as the physical problem is concerned, it is enough that it exists and is a function of position only. In that way, physics achieves great generality, since one may consider a wide variety of potential functions. As an example, we consider the famous problem of the vibrating string, first studied in 1714 by Brook Taylor (1685–1731)—famous for the Taylor series, for which he shares credit with Johann Bernoulli (1667–1748).

**Example 2.1.** The problem is to discuss the motion of a vibrating string. The best physical model to help our intuition is that of a violin string, clamped in two places and very long in comparison with the amplitude of its vibration. We shall assume the string at rest occupies an interval $(0, L)$ of the $x$-axis, and that its ends are fixed at all times. We are going to consider the motion of an infinitesimal portion of the string, which at rest lies between $x$ and $x+dx$. We wish to describe its height above the $x$-axis at time $t$ as a function $y(x,t)$. Now come all those linearizing and simplifying assumptions that were mentioned above. We assume that:

1. The density $\rho$ of the string, expressed in mass per unit length, is constant throughout the string and at all times. In other words, the change in density when the string is stretched is negligible.
2. The molecules of the string move only vertically; that is, their horizontal motion can be neglected.
3. The amplitude of the vibration is very small compared with the length of the string, so that the sine of the angle of inclination can be considered equal to the tangent of that angle, which is the slope $\partial y/\partial x$.
4. The tension $T$ in the string, which is the force the portions of the string to the left and right of a point exert on each other due to the stretching, is constant. That is, the small variation in tension as the string vibrates can be neglected.
5. There is no damping due to gravitation, internal friction and heating, or fluid resistance.

Under all these simplifying assumptions, we take the potential energy $E_p$ of the portion of the string between $x$ and $x+dx$ at rest to be due to the stretching of the string, that is it equals $-T$ multiplied by the change in length $d\ell$ of that portion, which is

$$d\ell = \left(\sqrt{1+\Big(\frac{\partial y}{\partial x}\Big)^2} - 1\right) dx\,.$$

Finally, we make one last linearizing assumption, replacing the square root in this last expression by the first two terms of its Maclaurin series:

$$\sqrt{1+\Big(\frac{\partial y}{\partial x}\Big)^2} = 1 + \frac{1}{2}\Big(\frac{\partial y}{\partial x}\Big)^2.$$

Thus, we have

$$E_p(x) = -T\,d\ell = -\frac{1}{2}T\Big(\frac{\partial y}{\partial x}\Big)^2.$$

The kinetic energy of this piece of the string is simply half of its mass, which is $\rho\,dx$ times the square of its velocity, and by our assumption the velocity is vertical. Thus

$$E_k(y') = \frac{1}{2}\rho\Big(\frac{\partial y}{\partial t}\Big)^2 dx\,.$$

Thus, canceling $-dx$ from both expressions for the energy, we find by the variational form of Newton's second law that the integral

$$\int_{t_0}^{t_1} \frac{1}{2}\rho\Big(\frac{\partial y}{\partial t}\Big)^2 + \frac{1}{2}T\Big(\frac{\partial y}{\partial x}\Big)^2\,dt$$

has a minimum. By Euler's equation, we have

$$\begin{aligned}\rho\frac{\partial^2 y}{\partial t^2} &= T\Big(\frac{\partial y}{\partial x}\Big)\frac{\partial}{\partial y}\Big(\frac{\partial y}{\partial x}\Big)\\ &= T\Big(\frac{\partial y}{\partial x}\Big)\frac{\partial}{\partial x}\Big(\frac{\partial y}{\partial x}\Big)\Big(\frac{\partial x}{\partial y}\Big)\\ &= T\frac{\partial^2 y}{\partial x^2}\,.\end{aligned}$$

In this way, we arrive at the classical linearized equation of the vibrating string:

$$\frac{\partial^2 y}{\partial t^2} = c^2\frac{\partial^2 y}{\partial x^2}\,.$$

Here the quantity $c^2 = T/\rho$ has the dimensions of force divided by linear density, which is to say the product of mass and length divided by time-squared, then divided by the quotient of mass and length. That means that $c$ is a velocity, in fact the velocity with which a transverse wave propagates along the string, as you will see in Problem 2.7 below.

Notice that, since the amplitude of the vibration is small compared with the length, the element of length along the string can be regarded as $dx$. In that case, the second derivative $\partial^2 y/\partial x^2$ represents the curvature at each point. As a result, the equation can be phrased succinctly by saying that the restoring force is proportional to the curvature. That was the principle from which Taylor originally derived it.

## 4. Problems

**Problem 2.1.** Prove the assertion in the text that if $f(t)$ is any continuous function on an interval $[a, b]$ such that

$$\int_a^b u(t)f(t)\,dt = 0$$

for all continuously differentiable functions $u(t)$ satisfying $u(a) = u(b) = 0$ or $u(a) = u'(a) = 0$, then $f(t) \equiv 0$ on that interval.

**Problem 2.2.** For a particle moving at constant speed along a geodesic of the surface $F(x, y, z) = f(x, y) - z = 0$, show that the third component of its acceleration is given by

$$z'' = \frac{\Delta}{1 + \big(\frac{\partial f}{\partial x}\big)^2 + \big(\frac{\partial f}{\partial y}\big)^2}\,.$$

In other words, the acceleration is $-k\,\nabla F(x, y, z)$, where

$$k = \frac{\Delta}{1 + \big(\frac{\partial f}{\partial x}\big)^2 + \big(\frac{\partial f}{\partial y}\big)^2}\,.$$

**Problem 2.3.** Show that the first fundamental form for the surface $F(x, y, z) = f(x, y) - z = 0$ is $E\,dx^2 + 2F\,dx\,dy + G\,dy^2$, where

$$\begin{aligned} E &= 1 + \Big(\frac{\partial f}{\partial x}\Big)^2, \\ F &= \frac{\partial f}{\partial x}\,\frac{\partial f}{\partial y}, \\ G &= 1 + \Big(\frac{\partial f}{\partial y}\Big)^2. \end{aligned}$$

**Problem 2.4.** Show that the element of surface area $dS$ on the surface $f(x, y) - z = 0$ is

$$dS = \sqrt{1 + \Big(\frac{\partial f}{\partial x}\Big)^2 + \Big(\frac{\partial f}{\partial y}\Big)^2}\,dy\,dx.$$

**Problem 2.5.** Show that the stars in the Ptolemaic frame of reference (see Section 12 of Chapter 1) are *not* moving along geodesics in the non-Euclidean geometry of the surface that was used to reconcile the motion with special relativity. (Therefore, they must have some acceleration tangential to the surface. In relativistic language, the space-time metric cannot be the flat one of the special theory of relativity.)

**Problem 2.6.** Show how to state the vibrating string problem in Newtonian language, using $F = ma$.

**Problem 2.7.** Show that one solution to the equation of the vibrating string, given by Jean le Rond d'Alembert (1717–1783), is

$$y(x, t) = \frac{f(x + ct) + f(x - ct)}{2},$$

where $y = f(x)$ is the initial configuration of the string at time $t = 0$, extended to all positive and negative values of $x$ as an odd function of period $2L$, and that this solution has initial velocity 0.

That is, the string is stretched into the shape given by this function and suddenly released. This solution shows that the (unique) solution to this initial-value problem when the ends of the string are clamped is the average of the two waves generated by moving the initial configuration right and left with velocity $c$.

**Problem 2.8.** Show that another way of solving the same problem, that is, given an initial configuration in the shape of the graph of $y = f(x)$ and initial velocity 0, is

$$y(x, t) = \sum_{n=1}^{\infty} c_n \sin\Big(\frac{n\pi x}{L}\Big)\cos(nct),$$

with the constants $c_n$ chosen so that

$$f(x) = \sum_{n=1}^{\infty} c_n \sin\Big(\frac{n\pi x}{L}\Big).$$

Make whatever assumptions as to smoothness, differentiability, and convergence of the series you find convenient. This solution was given by Daniel Bernoulli

(1700–1782), the son of Johann Bernoulli, a Swiss compatriot of Euler, and a colleague of Euler's for a while at the Russian Academy of Sciences during the 1730s.

## APPENDIX 3

# Point-Set Topology

The several varieties of topology—differential topology, algebraic topology, combinatorial topology, point-set topology—all have the common aim of developing techniques for dealing with continuous quantities such as the lines, planes, and solids of the very ancient Euclidean geometry. Topology differs from geometry in not requiring, at least in its basic structure, any metric concepts such as distance, area, volume, and congruence. These are quantitative concepts that require the use of real numbers, whereas general point-set topology is qualitative. The primary apparatus of the subject is a collection of sets called a *topology* and declared by fiat to be *open*. The fiat is subject only to four simple restrictions, listed below. From the topology, the basic trichotomy of interior, exterior, and boundary of a set is defined, and that leads to essentially the whole of the subject, which consists of the six "big C's": closure, connectedness, compactness, convergence, continuity, and category. Those concepts are the basic tools of analysis. They make it possible to prove things in a wide variety of contexts without invoking the more intricate properties of geometric figures. Once the core is developed, point-set topology does allow the metric camel to stick its nose under the topological tent-flap by introducing the concept of a metric space, in terms of which some important topologies can be defined. When that is done, a seventh "big C" arises: completeness.

Even though I am partial to the philosophy of Immanuel Kant, the present chapter is written in the spirit of Platonism, which seems to be the unconscious metaphysic behind modern set theory. The objects studied in point-set topology are said to "exist" as the result of certain axioms of set theory that begin with the symbol $\exists$ (read "there exists"). Thus, for example, there is an axiom asserting that if two sets $A$ and $B$ exist, a third set $A \cap B$ exists whose members are the elements that are members of both $A$ and $B$. When we write about these things, especially in proofs, we adopt the perspective of an all-seeing demigod perched atop some Olympus and able to perceive from that vantage point the Platonic realm in which mathematical objects, just like Plato's Forms or Ideas, live and move and have their being. This demigod is allowed to invent a symbol for any object that exists in that Platonic heaven and write sentences using that symbol that express truths about the object it represents. We shall write $A \cap B$ for the intersection of two sets $A$ and $B$, just as if we knew exactly what its elements are, even when we may not actually know *anything* about them, because we may not be able to describe uniquely (*name*) even one element of $A$ or $B$. For example, $A$ might be the set of finite unions of sets in the seventh level of the Borel hierarchy[1] and $B$ the set of finite intersections of such sets. Since no one has ever given an unambiguous

[1] The Borel hierarchy, named for Emile Borel (1871–1956), starts at level 0 with the class of sets that are either open or closed. Its first level consists of sets that are not themselves open or closed but are countable unions and countable intersections of open or closed sets. The second level

description of (*named*) any particular set above the fourth level of that hierarchy, we are flying through some very empyrean heights here. But the axioms of set theory imply that all these levels "exist" and that each of them contains sets not in any lower level.

When we say, "Let $A$ be the set of elements having property $P$," we are conjuring the set $A$ into "existence"—provided we do not violate any rules of set theory in our definition—by fiat.[2] Once we do so, we write about $A$ and its elements exactly as we would if they were physical objects that we could see. This convenient Olympian perspective can collide catastrophically with reality when it becomes necessary to describe or *name* a mathematical object in terms of the familiar points of Euclidean space or the real numbers. Even worse, when using proof by contradiction, we go to a second, even more remote, Platonic heaven made up of objects that *exist as ideas*, and where we can talk about the properties of objects that *would exist as ideas* if a certain hypothesis were true.[3] In that case also, we invent symbols for objects we have no way of actually finding, since it is possible that they do not "exist," even in the Pickwickian sense that the word has in the realm of set theory. To prove that such an object does not "exist," we try to show that the properties they would have to have lead to a contradiction. All that, as any applied mathematician or physicist or engineer would be quick to say, is a powerful expression of human *hubris*. We justify it Platonically by saying that these objects and truths are perceived by the mind rather than the senses, and that we can get our minds to think about them in exactly the way it thinks about the concrete physical objects that get into our minds through our senses. And, we add, the result of the study is a store of mental habits (concepts) that shape our thinking and clarify the essence of some very complicated mathematical phenomena. This extreme abstraction increases our power of analysis by stripping irrelevant concepts out of whatever problems we may be studying.

The study of manifolds, which is the main reason relativity theory needs topology, is part of differential topology, which is the most restrictive, least general, form of topology. It requires a concept of differentiability, whose roots go back to the beginnings of calculus, and therefore assumes the most elaborate structures of all the forms of topology. The least restrictive, most general form of topology, and the only other form we need to consider, is point-set topology, which grew out of real analysis alongside set theory. We begin with the most general form, very little of which is actually needed to study manifolds. Indeed, none at all would be needed if we wished to consider only manifolds that are subsets of some Euclidean space, perhaps of high dimension. But embedding a manifold in Euclidean space complicates some computations. If we study abstract manifolds instead, we can use simpler tools, at the cost of having to know some point-set topology.

## 1. Topological Spaces

A *topological space* is a non-empty set $X$, together with a distinguished collection $\mathfrak{T}$ of subsets of $X$, called a topology. The subsets in the topology are referred to as

consists of sets that are not in the zeroth or first level but are countable unions and countable intersections of these sets, and so on.

[2] In case the reader was unaware, *Fiat* ___ . is Latin for "Let ___ be."

[3] I like to think of the first Platonic heaven as the indicative realm of ideas and the second as the subjunctive realm of ideas.

*open sets*. Not every collection of subsets is a topology, only those that satisfy the following four axioms:

1. The empty set $\varnothing$ is an open set;
2. The set $X$ itself is an open set;
3. If $U$ and $V$ are both open sets, then $U \cap V$ is an open set;
4. If $U_\alpha$ is an open set for each index $\alpha$ in some index set $A$, then $\bigcup_{\alpha \in A} U_\alpha$ is an open set.

Examples of topologies abound, but by far the most important are subsets of a finite-dimensional Euclidean space $\mathbb{R}^n$, which has a concept of distance: If $\boldsymbol{x} = (x^1, \ldots, x^n)$ and $\boldsymbol{y} = (y^1, \ldots, y^n)$ the distance $d(\boldsymbol{x}, \boldsymbol{y})$ from $\boldsymbol{x}$ to $\boldsymbol{y}$ is defined by a generalized Pythagorean "metric" as

$$d(\boldsymbol{x}, \boldsymbol{y}) = \left((x^1 - y^1)^2 + \cdots + (x^n - y^n)^2\right)^{\frac{1}{2}}.$$

The points $\boldsymbol{x}$ and $\boldsymbol{y}$ should be visualized in the usual vector way, as arrows with tail at the origin $\mathbf{0}$ and heads at the points $P = (x^1, \ldots, x^n)$ and $Q = (y^1, \ldots, y^n)$. Then $d(\boldsymbol{x}, \boldsymbol{y})$ is thought of as the length of the line joining $P$ to $Q$. The reason for calling this particular notion of distance a Pythagorean metric is that if $\boldsymbol{x}$ and $\boldsymbol{y}$ are perpendicular, in the sense that their dot product $\boldsymbol{x} \cdot \boldsymbol{y} = x^1 y^1 + \cdots + x^n y^n$ equals 0, then

$$\left(d(\boldsymbol{x}, \boldsymbol{y})\right)^2 = \left(d(\mathbf{0}, \boldsymbol{x})\right)^2 + \left(d(\mathbf{0}, \boldsymbol{y})\right)^2.$$

**Example 3.1.** In Euclidean space, or indeed in any metric space $X$ with a notion of distance $d(P, Q)$ between any two distinct points that is a positive number, such that the distance from $P$ to $Q$ equals the distance from $Q$ to $P$ and is not greater than the sum of the distance from $P$ to a third point $R$ and the distance from $R$ to $Q$ (the triangle inequality), we can define an *open ball* $B_r(x)$ for any point $x \in X$ and any positive real number $r$ as follows:

$$B_r(x) = \{y \in X : d(x, y) < r\}.$$

We then say that a subset $U \subseteq X$ is open if for each $x \in U$, there is some positive number $r$ such that $B_r(x) \subseteq U$. Intuitively, $x$ is "surrounded" by points of $U$, and no path through $X$ that avoids $U$ can get closer to $x$ than the distance $r$.

It is trivial to verify that the open sets in $X$ form a topology, that is, satisfy the three required conditions. (The empty set $\varnothing$ satisfies the condition $B_r(x) \subseteq \varnothing$ for all $x \in \varnothing$ and $r > 0$ "vacuously" since there are no points $x \in \varnothing$.)

**Example 3.2.** At the opposite extreme from the beautiful manifolds that we shall be studying we can impose a trivial topology $\mathfrak{T}$ on any set $X$, the strongest possible topology, by simply declaring that *every* subset of $X$ is open. This topology is of limited use in geometry, since using it amounts to a decision to forego all topological considerations. (If topology is to be of any use in a given application, it needs to be a significant statement that a subset of $X$ is open, not a mere tautology.) Nevertheless, if we had any use for a "zero-dimensional" manifold, which can be thought of as a collection of isolated points, like a set of atoms in space, this would be the natural topology to impose (or not, since imposing it wouldn't achieve anything). Since a zero-dimensional manifold has no differentiable structure, we won't be using this topology.

This topology is called the *discrete topology* and it is defined by the so-called discrete metric whereby $d(P, Q) = 1$ if $P \neq Q$ and $d(P, P) = 0$ for all $P$ and $Q$ in $X$.

**Example 3.3.** Besides the "reasonable" topologies given by a metric, the definition of a topology is so general that many "pathological" examples exist. The minimum, or weakest topology, which being the opposite extreme from the discrete topology, is called the *indiscrete* topology (punning on the word *indiscreet*), consists of just the two sets $\varnothing$ and $X$. Again, there are no significant conclusions to be drawn by use of this topology. Anticipating the subsections that immediately follow, we merely mention that in this topology every subset is connected and compact, and the only continuous real-valued functions on such a space are those that are constant.

Less extreme topologies may still be too weak or too strong to make significant statements. One might, for instance, define a subset $E \subseteq X$ to be open if $E = \varnothing$ or if $X \setminus E$ is a finite set (the so-called *co-finite topology*). If $X$ itself is finite, this topology is the discrete topology, but not otherwise. The word *finite* here could be replaced by *countable* (the so-called *co-countable topology*), although again, if $X$ is itself countable, this topology is the discrete topology. We shall not have any use for such bizarre topologies as these, because, once again, they do not produce any non-trivial compact sets, connected sets, or continuous functions.

Thus, in defining a topology, the aim should be the *Goldilocks principle*: Not too strong, not too weak, just right. One way to keep the topology from being too weak is to require the following property, which is not very restrictive, but does make connectedness, compactness, and continuity significant.

**Definition 3.1.** A topological space $X$ is a *Hausdorff* space if for any two distinct points $x \in X$, $y \in X$, there exist disjoint open sets $U$ and $V$ (that is, $U \cap V = \varnothing$) such that $x \in U$ and $y \in V$.

A Hausdorff space, named after Felix Hausdorff (1868–1942), is sometimes said to be "$T_2$" since it satisfies the second in a series of increasingly strong "separation axioms," which we shall not elaborate, except to say that assuming only the existence of an open set $U$ such that $x \in U$ and $y \notin U$ is the $T_1$ property. The co-finite and co-countable topologies just mentioned are $T_1$-topologies, but not $T_2$-topologies. In his treatment of topology, Hausdorff made the $T_2$ property one of the basic axioms. We shall always assume that the manifolds we are going to study are Hausdorff spaces.

Intuitively, an open set is made up of points that are "close to one another." The intersection of all the open sets that contain a point $x$ ought to be only the point $x$ (and it is just that in a Hausdorff space, as one can trivially prove). This intuition comes from the example of metric spaces, where one can shrink down on a the point $x$ by taking balls $B_r(x) = \{y : d(x, y) < r$ for small positive values of $r$. The fundamental set-theoretic relation between points of a space $X$ and subsets of $X$ is a dichotomy: Given a point $x \in X$ and a subset $E \subseteq X$, either $x \in E$ or $x \notin E$, and precisely one of these two possibilities holds. If we have a topology, we can enlarge this dichotomy to a trichotomy. The possibilities are:

1. There is an open set $U$ such that $x \in U \subseteq E$;
2. There is an open set $V$ such that $x \in V$ and $V \cap E = \varnothing$, that is $V \subseteq X \setminus E$.
3. Every open set that $x$ belongs to contains both points of $E$ and points of $X \setminus E$.

The first possibility is a qualitative way of saying that $E$ "completely surrounds" the point $x$, that is, it contains not only $x$ but also all the points "sufficiently near" $x$. In that case, we shall say that $x$ belongs to the *interior* of $E$. Equivalently, we

shall say that $E$ is a *neighborhood* of $x$. It is trivial that a set is open if and only if it equals its interior. The second possibility says qualitatively that $x$ is "insulated from" $E$, in that it is "completely surrounded" by points not in $E$. In that case, $x$ belongs to the *exterior* of $E$. The third says that $E$ and its complement $X \setminus E$ "meet at" $x$, that is, $x$ can be "approached" from both inside $E$ and outside $E$. In that case, we shall say that $x$ belongs to the *boundary* of $E$. Thus, a set $E$ now partitions the space $X$ into three pairwise disjoint sets: the interior of $E$ (denoted $\operatorname{int}(E)$), the exterior of $E$ (denoted $\operatorname{ext}(E)$), and the boundary of $E$ (denoted $\partial E$ for reasons having to do with differentiation and integration). Obviously, $\operatorname{int}(E)$ consists of points in $E$ and $\operatorname{ext}(E)$ consists of points not in $E$, while $\partial E$, in general, contains both points in $E$ and points not in $E$. The sets $E$ for which $\partial E$ contains only points not in $E$ or only points in $E$ are precisely the open sets (the sets of the topology) and their complements, which are called *closed* sets.

**Example 3.4.** Let us start with the intuitive basis for these definitions in Euclidean space $\mathbb{R}^2$. Consider the set

$$E = \{(x^1, x^2) \in \mathbb{R}^2 \,:\, (x^1)^2 + (x^2)^2 < 1\}.$$

The set $E$ is the "open unit disk" in $\mathbb{R}^2$. It is open because if $(x^1, x^2) \in E$, then $(x^1)^2 + (x^2)^2 < 1$, and for the positive number $\delta = \sqrt{1 - (x^1)^2 - (x^2)^2}$, the ball $B_\delta(x^1, x^2)$, which consists of points $(y^1, y^2)$ such that $d\big((x^1, x^2), (y^1, y^2)\big) = \sqrt{\big((x^1 - y^1)^2 + (x^2 - y^2)^2\big)} < \delta$ is contained in $E$. Thus $E = \operatorname{int}(E)$, and so $E$ is open. By merely reversing the inequality, we see that the exterior of $E$ consists of the points $(x^1, x^2)$ such that $(x^1)^2 + (x^2)^2 > 1$. What is left—the unit circle, consisting of $(x^1, x^2)$ such that $(x^1)^2 + (x^2)^2 = 1$ is $\partial E$. Again this is easy to see, since if $(x^1)^2 + (x^2)^2 = 1$, then any ball $B_\delta(x^1, x^2)$ with $0 < \delta < 1$ contains the points $\big((1 - \delta/2)x^2, (1 - \delta/2)x^2\big)$ and $\big((1 + \delta/2)x^1, (1 + \delta/2)x^2\big)\big)$, the first of which is in $E$ and the second of which is not in $E$.

**Example 3.5.** If we modify the set $E$ in the preceding example by including part or all of the boundary circle as part of $E$ and leaving the remaining part of the circle in $\mathbb{R}^2 \setminus E$, the new set $\tilde{E}$ so defined has exactly the same interior, exterior, and boundary as the set $E$. Since it does not equal its interior it is not open. If we include the whole boundary as part of $E$, then $X \setminus E = \operatorname{ext}(E)$, which is an open set.

**Example 3.6.** For a less intuitive example, still in $\mathbb{R}^2$, consider the set of points both of whose coordinates are rational. This set has an empty interior and an empty exterior. Its boundary is all of $\mathbb{R}^2$. We don't normally think of boundaries that way, but that is because we are usually thinking about figures that resemble the familiar ones from Euclidean geometry. For those familiar figures, the interior, exterior, and boundary are what we would expect. If we look at more esoteric examples, we must fall back on the definitions. Still, the idea of the boundary of a set $E$ as the sort of no-man's-land between $E$ and $X \setminus E$, "undefended" by both $E$ and $X \setminus E$, is a useful picture to keep in mind.

The formal consequences of this trichotomy turn out to be very extensive, and we now turn to the task of developing them.

**1.1. Open and closed sets.** The complement of an open set is naturally called a *closed* set. Thus, $X$ and $\varnothing$, being complementary open sets, are also closed sets. It

is important to keep this example in mind, since it shows that *open* and *closed* are not opposites. A set may be both simultaneously. In general, a set is neither, one exception being the discrete topology, in which every set is both open and closed. In proving that a set is open, then, it does not help to prove that it isn't closed, since it may be neither. It also may not even be possible to prove that it isn't closed, since it may be both.

We begin our study of point-set topology by recasting the informal definitions of interior, exterior and boundary given above. The equivalence of the new definitions to the old should be obvious.

**Definition 3.2.** If $X$ is a topological space and $E$ is any subset of $X$, the union of all open sets $U$ contained in $E$ is an open set, called the *interior* of $E$, and denoted $\operatorname{int}(E)$. The interior of $E$ is obviously open and is the largest open set contained in $E$. If $x \in \operatorname{int}(E)$, we also say that $E$ is a *neighborhood* of $x$. In other words, $E$ is a neighborhood of $x$ ($x$ is an interior point of $E$) if and only if there is an open set $U$ such that $x \in U \subseteq E$.

**Definition 3.3.** The interior of the complement of a set $E$ ($\operatorname{int}(X \setminus E)$) is called the *exterior* of $E$ and denoted $\operatorname{ext}(E)$. The set $\operatorname{ext}(E)$ is the largest open set disjoint from $E$. Points that belong to neither the interior of $E$ nor the exterior of $E$ are called *boundary* points of $E$. The boundary of $E$ is denoted $\partial E$. Being the complement of the open set $\operatorname{int}(E) \cup \operatorname{ext}(E)$, the boundary $\partial E$ is always a closed set.

Two obvious facts are left as exercises to the reader to prove: (1) $\partial E = \partial(X \setminus E)$ (Problem 3.1); (2) a set $E$ is closed if and only if $\partial E \subseteq E$ (Problem 3.2).

**Definition 3.4.** The *closure* of the set $E$ is $E \cup \partial E$ and is denoted $E^c$.

Thus, each subset of $E$ determines the topological trichotomy given above for points of $X$. For each point $x \in X$ exactly one of the following is true:

1. $x \in \operatorname{int}(E)$;
2. $x \in \operatorname{ext}(E)$;
3. $x \in \partial E = \partial(X \setminus E)$.

**1.2. Connectedness.** The manifolds we study will be assumed to be connected. This assumption, as we shall see, has two advantages: First, it ensures that the dimension of the manifold is the same at every point. Second, it implies that any two points on the manifold can be connected by a continuous curve. All of these concepts, of course, have first to be defined.

**Definition 3.5.** A topological space $X$ is *connected* if the only sets that are both open and closed are the trivial ones. In other words, only these two sets have empty boundary.

More generally, a subset $E \subseteq X$ is connected if and only if the only subsets of $E$ that are both *open relative to* $E$ and *closed relative to* $E$, that is, for which there exist an open set $U \subseteq X$ and a closed set $C \subseteq X$ such that $E \cap U = E \cap C$, are $\varnothing$ and $E$ itself. (If you take $U = X$ and $C = \varnothing$, you see that $E$ is open relative to $E$ and $\varnothing$ is closed relative to $E$. If you take $U = \varnothing$ and $E = X$, you see that $E$ is closed and $\varnothing$ is open relative to $E$. In effect, these considerations amount to imposing a topology $\mathfrak{T}_E$ on $E$, whereby a subset $F$ of $E$ is open if and only if

$F = E \cap U$ for a set $U$ that is open in $X$. The topology $\mathfrak{T}_E$ on $E$ is said to be *inherited from* or *induced by* the topology $\mathfrak{T}$ on $X$.

It is generally easier to say what it means for $E$ *not* to be connected. A subset $E$ of $X$ is *not* connected if there exist two open sets $U$ and $V$ of $X$ such that $U \cap E \neq \varnothing \neq V \cap E$, $E \subseteq U \cup V$, and $U \cap V \cap E = \varnothing$. To prove that a set is connected, we often proceed by contradiction, assuming a "disconnection" determined by two open sets $U$ and $V$, as above, and then showing that it is impossible. For example, let us prove a simple theorem:

**Theorem 3.1.** *If $E$ is a connected subset of a topological space $X$ and $E \subseteq F \subseteq E^c$, then $F$ is connected.*

PROOF. Let $E$ and $F$ be as in the statement of the theorem, and assume that $F$ is not connected. Then there are two open sets $U$ and $V$ such that $F \subseteq U \cup V$, $F \cap U \cap V = \varnothing$, $U \cap F \neq \varnothing$, and $V \cap F \neq \varnothing$, that is $x_1 \in U \cap F$ and $x_2 \in V \cap F$. Because $U \cap V \cap F = \varnothing$, $x_1$ and $x_2$ are different points. They do not both belong to $E$, since if they did, $E$ would be disconnected, contrary to hypothesis. Therefore, at least one of them, say $x_1$, belongs to $\partial E$ but not to $E$. Since $U$ is an open set containing $x_1$, it contains a point of $E$, say $x_3$. (Otherwise, $x_1$ would belong to $\operatorname{ext}(E)$.) It then follows that $V$ contains no points of $E$, since otherwise, once again, $U$ and $V$ would disconnect $E$. But that means that $x_2$ also belongs to the boundary of $E$, and hence that $V$ *does* contain a point $x_4$ in $E$. Yet again, $U$ and $V$ must disconnect $E$. Having retreated before the assaults of logic at every point, we have finally reached a cul-de-sac and are caught in an unavoidable contradiction. We conclude that no such disconnection of $F$ as we assumed can exist. □

In connection with the two motivations for considering the notion of connectedness, we can prove a simple proposition that is occasionally useful. In order to state it, we need the concept of an *equivalence* relation $R$ on a set $X$. This is a relation that either does or does not hold between any two points of $X$. If it holds between $x$ and $y$, we write $x\,R\,y$.

**Definition 3.6.** A binary relation $R$ is an equivalence relation if (and only if) it has the following three properties: (1) $x\,R\,x$ for all $x \in X$ (the reflexive property); (2) for all $x \in X$ and $y \in X$, if $x\,R\,y$, then $y\,R\,x$ (the symmetric property); (3) for all $x \in X$, $y \in X$, $z \in X$, if $x\,R\,y$ and $y\,R\,z$, then $x\,R\,z$ (transitive property).

The strongest possible equivalence relation on $X$ is identity, that is to say $x\,R\,y$ means that $x$ and $y$ are the same point of $X$. The weakest possible equivalence relation on $X$ is the vacuous one, that is to say, $x\,R\,y$ means $x \in X$ and $y \in X$ (in other words, $x\,R\,y$ means nothing more than was already assumed about $x$ and $y$). In between these, there are enormous numbers of more or less useful equivalence relations. For example, if $X$ is the set of positive integers and $m$ is a fixed positive integer, the relation $x\,R\,y$ might mean that $x - y$ is divisible by $m$ (equality modulo $m$). As a second example, if $f : X \to Y$ is any mapping from $X$ into a set $Y$, we might define $x\,R\,y$ to mean that $f(x) = f(y)$. As a third example, if $X$ is the set of all mappings from a set $A$ into a second set $B$, we might define $f\,R\,g$ to mean that $f(a) = g(a)$ for all $a \in A$ with at most a finite number of exceptions.

The main fact about equivalence relations is the following:

**Theorem 3.2.** *If $R$ is an equivalence relation on $X$, then $R$ partitions $X$. That is, $R$ determines a unique collection of subsets $E_\alpha \subseteq X$ such that* (1) *every point*

*$x \in X$ belongs to exactly one of the sets $E_\alpha$ and* (2) *for all $x \in X$, $y \in X$, the relation $x\,R\,y$ holds if and only if $x$ and $y$ belong to the same set $E_\alpha$.*

The sets $E_\alpha$ are called *equivalence classes* of the equivalence relation $R$.

PROOF. For each $x \in X$, let $E_x$ be the set of $y \in X$, such that $x\,R\,y$. Then $x \in E_x$ by the reflexive property, and therefore the classes $E_x$ cover $X$, that is, every point $x$ belongs to one of the sets. If $z \in E_x \cap E_y$ for some point $z \in X$, we claim that $E_x = E_y$, so that each point belongs to exactly one of these sets. Indeed, $z \in E_x$ means $x\,R\,z$, and $z \in E_y$ means $y\,R\,z$. By the symmetric property, we then find that $z\,R\,y$, and so by the transitive property $x\,R\,y$, and therefore $y \in E_x$. But then, by the transitive property, $y\,R\,w$ implies $x\,R\,w$, and so $E_y \subseteq E_x$. The same reasoning shows that $E_x \subseteq E_y$, and therefore, if $E_x \cap E_y \neq \varnothing$, then $E_x = E_y$. Thus, two equivalence classes are either coincident or disjoint, as asserted. □

This theorem has an obvious converse. That is, every partition of $X$ into pairwise disjoint subsets determines an equivalence relation $R$. Simply define $x\,R\,y$ to mean that $x$ and $y$ belong to the same subset of the partition.

If $X$ is a topological space, we can define a partition of $X$ by defining a set $E_x$ for each $x \in X$ to be the union of all connected subsets of $X$ to which $x$ belongs. Obviously the singleton set $\{x\}$ is a connected set, so that $x \in E_x$. Thus the sets $E_x$ cover $X$. We claim that $E_x$ is itself a connected set, and hence is the unique maximal connected set containing $x$ (or any other point that belongs to $E_x$). Indeed, suppose that $U$ and $V$ are open sets in $X$ such that $E_x \subseteq U \cup V$, and $E_x \cap U \cap V = \varnothing$. We claim that $U$ and $V$ do not disconnect $E_x$. Without loss of generality, assume $x \in U$. We are claiming that $E_x \subseteq U$. Indeed, let $y \in E_x$. By definition of $E_x$, there is a connected set $F \subseteq E_x$ such that $x \in F$ and $y \in F$. It then follows that $y \in U$ also, since otherwise $U$ and $V$ would disconnect $F$. Thus, $E_x \cap V = \varnothing$, and the sets $U$ and $V$ do *not* disconnect $E_x$. Thus, there is no way $E_x$ can be disconnected, and it is therefore connected.

We next show that $E_x \cup E_y$ is connected if $E_x \cap E_y$ is non-empty. To that end, assume $z \in E_x \cap E_y$. Let $U$ and $V$ be open sets such that $E_x \cup E_y \subseteq U \cup V$ and $(E_x \cup E_y) \cap U \cap V = \varnothing$. Without loss of generality, assume $z \in U$. Then, as just shown, $E_x \subseteq U$ (because if $x \in V$, then $E_x \subseteq V$, which is not possible if $z \in U$). And similarly $E_y \subseteq U$. Therefore $E_x \cup E_y \subseteq U$ also, and so $U$ and $V$ do not disconnect $E_x \cup E_y$, which therefore has no disconnection and hence is connected. By the maximality implied by the definition of $E_x$, it follows that $E_x \cup E_y = E_x$, that is, $E_y \subseteq E_x$. And similarly, $E_x \subseteq E_y$, whence $E_x = E_y$. Thus we have a partition of $X$, and hence an implicit equivalence relation on $X$. For this relation, these equivalence classes, which are *maximal* connected subsets of $X$, are called the *components* of $X$.

**Theorem 3.3.** *The components of $X$ are connected subsets of $X$ and are maximal in the sense that if $E$ is a component of $X$ and $F \setminus E \neq \varnothing$, then $E \cup F$ is not connected. Components are closed subsets of $X$. If $X$ has only finitely many components, then the components are also open subsets of $X$.*

PROOF. The first statement has already been proved. The fact that components are closed follows from the maximality and the earlier proposition that if $E$ is connected and $E \subseteq F \subseteq E^c$, then $F$ is connected. The third follows from the fact that the complement of a component is the union of the other components. If there

are only finitely many of them, that union is closed, and hence the component is open. □

We note the example of the rational numbers $\mathbb{Q}$, for which the components are one-point sets. For if $a$ and $b$ are rational numbers with $a < b$, we can take $A$ to be the set of positive rational numbers $r$ such that $2r^2 - 4ar + (a^2 + 2ab - b^2) < 0$ and $V$ the set of $r$ such that $2r^2 - 4ar + (a^2 + 2ab - b^2) > 0$. It is easy to see that $a \in U$ and $b \in V$, and that $U \cap V = 0$. The fact that $U \cup V = \mathbb{Q}$ follows, since there are no rational numbers $r$ for which $2r^2 - 4ar + (a^2 + 2ab - b^2) = 0$. (The two real numbers that satisfy this quadratic equation are the irrational numbers $a \pm (b - a)/\sqrt{2}$.) A space like $\mathbb{Q}$ in which the components are singleton-point sets is said to be *totally disconnected*. Notice that the components of $\mathbb{Q}$ are *not* open sets. Components *may be*, and often are, both open and closed; but in general they *need not be* open.

Here is the simple proposition mentioned above, the one that makes it possible to prove some useful and interesting facts about connected spaces.

**Theorem 3.4.** *Let $X$ be a connected topological space and $R$ an equivalence relation on $X$ for which the equivalence classes are open sets. Then there is only one equivalence class, namely $X$ itself.*

Proof. Since the complement of an equivalence class is the union of the other equivalence classes, and equivalence classes are open, it follows that equivalence classes are also closed sets. Because $X$ is connected, there can be only two such sets, namely $\varnothing$ and $X$ itself. Since an equivalence class, by definition, is not empty, the sole equivalence class must be $X$ itself. □

One application of this theorem, given the Brouwer invariance-of-domain theorem,[4] is an easy proof that the dimension of a connected manifold must be the same at all points. Another application is a simple proof that that a connected manifold is path-connected.

**1.3. Compactness.** A second topological concept of great importance in analysis, and one we cannot do without, is the concept of compactness. The definition—and the reason for the name—as given in most textbooks, is rather obscure.

**Definition 3.7.** A subset $E$ of a topological space $X$ is *compact* if for any collection of open sets $U_\alpha$ such that $E \subseteq \cup U_\alpha)$, where $\alpha$ ranges over some index set $A$, there is a *finite* set of indices $\alpha_1, \ldots, \alpha_n$ such that $E \subseteq U_{\alpha_1} \cup \cdots \cup U_{\alpha_n}$.

This definition is expressed colloquially by saying that every open covering of $A$ contains a finite subcovering. A typical use of this definition is to show that a continuous real-valued function whose values are positive at every point of a compact set is actually "bounded away" from zero on that set. That is, not only do the values never reach zero, they do not even approach arbitrarily close to zero; they are "fenced off" by some positive minimum value $\varepsilon_0$.

Compactness can be made more intuitive, and the motivation for the name clearer, by introducing the concept of a *point of accumulation* of a set $E$. A point $x$ is a point of accumulation of $E$ if (and only if) every neighborhood of $x$ contains a

[4] This theorem, named after its discoverer Luitzen Egbertus Jan Brouwer (1881–1966) asserts that an open set in
*mathbbR*$^m$ is *not* homeomorphic to an open set in $\mathbb{R}^n$ if $m \neq n$.

point of $E$ *different from* $x$ *itself*. This idea is sometimes expressed by saying that every *deleted neighborhood of* $x$ (a set $N \setminus \{x\}$, where $N$ is a neighborhood of $x$) contains a point of $E$. The set of all points of accumulation of $E$ is denoted $E'$ and is called the *derived set* of $E$.

It is very easy to see that $E' \setminus E = \partial E \setminus E$, that is, a point not in $E$ is a point of accumulation of $E$ if and only if it is a boundary point of $E$. It follows in particular that $E^c = E \cup E'$. An interior point of $E$ can also be a point of accumulation, and in fact *is* a point of accumulation if $X$ is a manifold. Intuitively, points of accumulation are places where the set $E$ "bunches up" or "clusters." With that intuition in mind, we can explain the use of the term *compact*.

**Theorem 3.5.** *Let $E$ be a compact subset of $X$, and $F$ any infinite subset of $E$. Then $F' \cap E$ is non-empty.*

This theorem says intuitively that an infinite subset $F$ cannot be fitted into a compact set $E$ without getting "squeezed" at some point of $E$; hence the description of such a set as *compact*. In other words, a compact set cannot "go off to infinity" or even "go a finite distance" in infinitely many different dimensions. (It is a theorem of analysis that in an infinite-dimensional space with a metric given by a norm, a closed ball is *not* compact; but in a finite-dimensional space, it is.)

PROOF. Suppose $F$ is a subset of $E$ without any points of accumulation in $E$. We need to prove that $F$ is a finite set. If $x$ is any point of $E$, then $x$ is not a point of accumulation of $F$, and hence there is an open set $U_x$ such that $x \in U_x$, and $U_x \setminus \{x\}$ contains no points of $F$. Obviously, the sets $U_x$ cover $E$ as $x$ ranges over $E$. Since $E$ is compact, some finite number of them cover $E$, that is, $E \subseteq U_{x_1} \cup \cdots \cup U_{x_n}$ for some finite set of points $x_1, \ldots, x_n$ in $E$. But then, since $F \cap U_{x_i} \subseteq \{x_i\}$, it follows that $F = F \cap E \subseteq F \cap (U_{x_1} \cup \cdots \cup U_{x_n}) = \big(F \cap U_{x_1}\big) \cup \cdots \cup \big(F \cap U_{x_n}\big) \subseteq \{x_1, \ldots, x_n\}$, and so $F$ is a finite set. □

The converse of this theorem holds if $X$ is a complete metric space, such as Euclidean space $\mathbb{R}^n$. The statement that a bounded infinite subset of $\mathbb{R}^n$ has a point of accumulation is called the *Bolzano–Weierstrass theorem*, named after Bernhard Bolzano (1781–1848) and Karl Weierstrass (1815–1897).

**Theorem 3.6.** *A closed subset of a compact set is itself compact.*

PROOF. Let $E$ be a compact set and $F$ a closed subset of $E$. Suppose $F$ is covered by a collection of open sets $U_\alpha$. Then $E$ is covered by these sets together with the single open set $X \setminus F$. Hence some finite subcollection covers $E$, say $X \setminus F$, $U_{\alpha_1}, \ldots$, $U_{\alpha_n}$. These sets also cover $F$, since $F$ is contained in $E$. Since the first of them contains no points of $F$, it follows that the others cover $F$, and these sets therefore constitute a finite subcovering of $F$. □

Among the reasons for preferring Hausdorff spaces to all others is the following simple proposition.

**Theorem 3.7.** *A compact subset of a Hausdorff space is a closed set.*

PROOF. Let $X$ be a Hausdorff space and $E$ a compact subset of $X$. We need to show that $X \setminus E$ is open. To that end, fix a point $x \notin E$. For each $y \in E$, since obviously $y \neq x$, there are open sets $U_{xy}$ and $V_{xy}$ such that $x \in U_{xy}$, $y \in V_{xy}$, and

$U_{xy} \cap V_{xy} = \varnothing$. The open sets $V_{xy}$ cover $E$, and so some finite number of them, say $V_{xy_1}, \ldots, V_{xy_n}$ also cover $E$. The open set $U = U_{xy_1} \cap \cdots \cap U_{xy_n}$ contains no points of $V_{xy_1} \cup \cdots \cup V_{xy_n}$, and hence no points of $E$. That is, $x \in U \subseteq X \setminus E$, so that $x \in \operatorname{int}(X \setminus E)$. Since $x$ was an arbitrary point of $X \setminus E$, it follows that this set equals its interior and hence is an open set. □

In the important case of the space $X = \mathbb{R}^n$, it is obvious that a compact set $E$ is bounded, that is, contained in some ball $B_r(\mathbf{0})$ of finite radius $r$. For the open sets $B_n(\mathbf{0})$, $n = 1, 2, \ldots$ cover $E$ (since they cover all of $\mathbb{R}^n$!), and hence some finite number of them will cover $E$, that is, $E \subseteq B_n(\mathbf{0})$, where $n$ is the index of the largest ball in the finite collection that covers $E$. (The ball $B_n(\mathbf{0})$ covers $E$ all by itself, and the others are not needed for the cover.) Thus a compact subset of $\mathbb{R}^n$ is closed and bounded, and in fact this is true in any metric space. The important converse of this result for $\mathbb{R}^n$ is known as the *Heine–Borel theorem*:

**Theorem 3.8.** *A closed and bounded subset of $\mathbb{R}^n$ is compact.*

Proof. Since a closed subset of a compact set is compact, it suffices to prove that the $n$-dimensional cube of side $2s$, denoted $C^n$ and defined for a fixed $s > 0$ as

$$C^n = \{\boldsymbol{x} \in \mathbb{R}^n : -s \le x^1 \le s, \ldots, -s \le x^n \le s\},$$

is compact. We proceed using proof by contradiction.

Suppose the collection of open sets $U_\alpha$ cover $C^n$ but no finite subcollection does. Then one of the $2^n$ subcubes of side $s$, which we label $C_1^n$ and define as

$$C_1^n = \{\boldsymbol{x} \in \mathbb{R}^n : a_1^1 \le x^1 \le b_1^1, \ldots, a_1^n \le x^n \le b_1^n\},$$

where each $a_1^i$ is either $-s$ or 0 and $b_1^i = a_1^i + s$, cannot be covered by a finite subcollection of the $U_\alpha$. If all of these subcubes could be so covered, then the whole $n$-dimensional cube $C^n$ could be covered by a finite subcollection, contrary to hypothesis.

Next, one of the $2^n$ subcubes of side $s/2$ contained in $C_1^n$, a subcube that we label $C_2^n$ and define as

$$C_2^n = \{\boldsymbol{x} \in \mathbb{R}^n : a_2^1 \le x^1 \le b_2^1, \ldots, a_2^n \le x^n \le b_2^n\},$$

where each $a_2^i$ is either $a_1^i$ or $a_1^i + s/2$ and $b_2^i = a_2^i + s/2$, cannot be covered by a finite subcollection of the $U_\alpha$, for the same reason as before.

In this way, we get a nested sequence of closed $n$-dimensional cubes $C_k^n$ defined as

$$C_k^n = \{\boldsymbol{x} \in \mathbb{R}^n : a_k^1 \le x^1 \le b_k^1, \ldots, a_k^n \le x^n \le k_2^n\},$$

where each $a_k^i$ is either $a_{k-1}^i$ or $a_{k-1}^i + s/2^{k-1}$ and $b_k^i = a_k^i + s/2^{k-1}$, and no finite subcollection of the sets $U_\alpha$ covers $C_k^n$.

Now for each $i$, the closed intervals $[a_k^i, b_k^i]$ are nested, that is $a_1^i \le a_2^i \le \cdots \le a_k^i \le a_{k+1}^i < b_{k+1}^i \le b_k^i \le \cdots \le b_2^i \le b_1^i$. Thus, each $a_k^i$ is a lower bound for the whole set of points $b_m^i$, and each $b_k^i$ is an upper bound for the whole set of points $a_m^i$. If $L^i$ is the least upper bound of the $a_m^i$ and $M^i$ is the greatest lower bound of the $b_m^i$, we thus have $L^i \le M^i$. But since $b_k^i - a_k^i = s/2^{k-1}$, we have $M^i - L^i \le s/2^{k-1}$ for all $k$ and thus $L^i = M^i$. Let $\boldsymbol{x}_0 = (L^1, \ldots, L^n) = (M^1, \ldots M^n)$. Then $\boldsymbol{x}_0 \in C^n$, and hence there is some open set $U_{\alpha_0}$ in the collection containing $\boldsymbol{x}_0$. Being an open set $U_{\alpha_0}$ contains the ball $B_{r_0}(\boldsymbol{x}_0)$ for some $r_0 > 0$. If we choose $k$ so large that $2^{k-1} > s\sqrt{n}/r_0$ (that is, $r_0 > s\sqrt{n}/2^{k-1}$, we then have $a_k^i \le x_0^i \le b_k^i = a_k^i + s/2^{k-1}$,

and so $|x_k^i - t| \le s/2^{k-1}$ for all $t \in [a_k^i, b_k^i]$. Then if $\boldsymbol{x} \in C_k^n$ we have $\boldsymbol{x} = (t^1, \dots, t^n)$ for some numbers $t^i \in [a_k^i, b_k^i]$. Therefore

$$|\boldsymbol{x} - \boldsymbol{x}_0| = \big((x_0^1 - t^1)^2 + \cdots + (x_0^n - t^n)^2\big)^{\frac{1}{2}} \le \big(s^2 n/2^{2k-2}\big)^{\frac{1}{2}} = s\sqrt{n}/2^{k-1} < r_0\,.$$

This means that $C_k^n \subseteq B_{r_0}(\boldsymbol{x}_0) \subseteq U_{\alpha_0}$. But that in turn says that $C_k^n$ is covered by *one* set among the $U_\alpha$, and 1 is certainly a finite number! We have now reached the contradiction we were seeking, and we conclude that $C^n$ is compact, hence that every closed and bounded subset of $\mathbb{R}^n$ is compact. □

One last result on compactness will be useful below.

**Theorem 3.9.** *Let $A$ and $B$ be disjoint compact subsets of a Hausdorff space. Then there are disjoint open sets $U$ and $V$ with $A \subseteq U$ and $B \subseteq V$.*

The proof of this theorem uses the same principle used above to show that a compact subset of a Hausdorff space is closed.

Proof. For each point $x \in A$ and each point $y \in B$, there are disjoint open sets $U_{xy}$ and $V_{xy}$ such that $x \in U_{xy}$ and $y \in V_{xy}$. The open sets $U_{xy}$ cover $A$ as $x$ ranges over $A$. Hence some finite subcollection of them covers $A$, that is, $A \subseteq U_{x_1 y} \cup \cdots \cup U_{x_m y} = U_y$. The finite intersection $V_{x_1 y} \cap \cdots \cap V_{x_m y} = V_y$ is then an open set containing $y$ and disjoint from $U_y$. The open sets $V_y$ cover $B$ as $y$ ranges over $B$, and hence some finite subcollection of them covers $B$, that is, $B \subseteq V_{y_1} \cup \cdots \cup V_{y_n} = V$. The finite intersection $U_{y_1} \cap \cdots \cap U_{y_n} = U$ is then an open set containing $A$ and disjoint from the open set $V$, which contains $B$. □

## 2. Abstract Functions

If $f$ is any function having as domain a topological space $X$ and range (co-domain) a topological space $Y$, and $E$ is any set whatsoever, the *image* of $E$ under $f$, denoted $f(E)$, is defined to be the set of all values $f(x)$ for which $x \in E \cap X$. The image $f(E)$ is empty unless $E \cap X \neq \varnothing$, and obviously non-empty when the latter condition holds.

Next: If $f$ is any function, the *inverse image* of a set $E$ is denoted $f^{-1}(E)$ and consists of all points $x$ in the domain of $f$ such that $f(x) \in E$. This inverse image will be empty unless $E$ intersects the range $Y$ of $f$. Even if $E \cap Y \neq \varnothing$, it should not be thought that the set $f^{-1}(E)$ represents the image of $E \cap Y$ under a new function $f^{-1} : E \cap Y \to X$ that is the inverse of the function $f$. That function exists if and only if $f$ is *one-to-one*, that is, $x \neq y \Rightarrow f(x) \neq f(y)$ for any $x \in X$ and $y \in X$, in which case it is defined by saying that $f^{-1}(y)$ is the (unique) point $x \in X$ such that $f(x) = y$. The equations $f(x) = y$ and $f^{-1}(y) = x$ are equivalent in that case. Even if $f^{-1}$ is not defined as a function, the *set* $f^{-1}(E)$ always has meaning, though it may very possibly be the empty set $\varnothing$. When the inverse function $f^{-1}$ is defined, the notation $f^{-1}(E)$ might appear to be ambiguous, defining either the inverse image of $E$ under the mapping $f$ or the image of $E$ under the mapping $f^{-1}$. It is easy to see, however, that these two sets are the same in that case.

Finally, for any two functions $f_1$ and $f_2$ whatsoever, having domains $X_1$ and $X_2$ and ranges $Y_1$ and $Y_2$, if the set $X_{21} = f_1^{-1}(X_2) \cap X_1$ is non-empty, we can take it to be the domain of a function $f_2 \circ f_1$. By definition, $f_2 \circ f_1(x) = f_2\big(f_1(x)\big)$ for all $x \in X_{21}$. The range of $f_2 \circ f_1$ is the subset of $Y_2$ consisting of all the points $y = f_2(z)$ for which $z \in X_2 \cap Y_1$. The composition $f_2 \circ f_1$ is defined as a function only under

these special conditions. For the sake of brevity, however, we may wish to make statements about this composition even when these special conditions do not hold. In that case (the case when $X_{21} = \varnothing$), we make the bizarre convention that $f_2 \circ f_1$ is a *nonexistent function*, and that any assertion whatever about any function whatever is automatically true of all nonexistent functions. This convention is analogous to the logical convention whereby a proposition $P(x) \Rightarrow Q(x)$ is regarded as "vacuously true" when the set of $x$ for which $P(x)$ is defined is empty. Thus, for example, a nonexistent function $f_2 \circ f_1$ is infinitely differentiable if $f_1$ and $f_2$ are. This convention saves us the trouble of writing "if the composition exists" countless times. When verifying or proving a proposition about compositions, obviously, we can ignore non-existent compositions, and assume at the outset that $X_{21}$ is not empty.

**2.1. Continuity.** Point-set topology makes it easy to say what we mean by a continuous function:

**Definition 3.8.** A function $f : X \to Y$ is *continuous at a point* $x_0 \in X$ if (and only if) the inverse image of every neighborhood of $f(x_0)$ is a neighborhood of $x_0$.

In colloquial language, this definition says that all points "near" $x_0$ map to points "near" $f(x_0)$. (Other points not "near" $x_0$ may also map to points "near" $f(x_0)$, but the important thing is that *if* a point $x$ is "near" $x_0$, then $f(x)$ is "near" $f(x_0)$.)

**Theorem 3.10.** *A function $f : X \to Y$ is continuous at all points of $X$ if and only if $f^{-1}(U)$ is an open subset of $X$ for all open sets $U \subseteq Y$.*

PROOF. If $f$ is continuous at every point and $U$ is an open subset of $Y$, then $U$ is a neighborhood of each of its points, and therefore $f^{-1}(U)$ is a neighborhood of each of its points. That says precisely that every point of $f^{-1}(U)$ is an interior point, and hence that $f^{-1}(U)$ is open. The converse statement is trivial. □

**Corollary 3.1.** *A function $f : X \to Y$ is continuous at all points of $X$ if and only if $f^{-1}(C)$ is a closed subset of $X$ for all closed sets $C \subseteq Y$.*

PROOF. Suppose $f$ is continuous at all points of $X$. If $C$ is a closed subset of $Y$, then $U = Y \setminus C$ is an open subset of $Y$. Hence $f^{-1}(U)$ is an open subset of $X$, and $X \setminus f^{-1}(U) = f^{-1}(Y \setminus U) = f^{-1}(C)$ is a closed subset of $X$.

Conversely, if $f^{-1}(C)$ is a closed subset of $X$ whenever $C$ is a closed subset of $Y$, then $f^{-1}(U) = X \setminus f^{-1}(Y \setminus U)$ is an open subset of $X$ whenever $U$ is an open subset of $Y$. □

**Corollary 3.2.** *If $f : X \to Y$ is continuous at all points, and $E \subseteq X$, then $f\big(E^c\big) \subseteq \big(f(E)\big)^c$.*

PROOF. Assume $f$ is continuous at all points. The set $C = \big(f(E)\big)^c$ is closed, so that $f^{-1}(C)$ is closed. Since it contains $E$, it also contains $E^c$, which is the smallest closed set containing $E$. That says precisely that $f(E^c) \subseteq \big(f(E)\big)^c$. □

This last corollary can be phrased succinctly by saying that *the image of the closure of a set under a continuous mapping is contained in the closure of the image of the set.*

The important uses of connectedness and compactness are contained in the following theorems:

**Theorem 3.11.** *Let $X$ and $Y$ be topological spaces and $f : X \to Y$ a continuous mapping. If $E$ is a connected subset of $X$, then $f(E)$ is a connected subset of $Y$.*

PROOF. It is easy to see that if $U$ and $V$ are open sets that disconnect $f(E)$, then their inverse images $f^{-1}(U)$ and $f^{-1}(V)$ are open sets that disconnect $E$. □

**Theorem 3.12.** *Let $X$ and $Y$ be topological spaces and $f : X \to Y$ a continuous mapping. If $E$ is a compact subset of $X$, then $f(E)$ is a compact subset of $Y$.*

PROOF. It is easy to see that if $\{U_\alpha\}$ is an open covering of $f(E)$ having no finite subcovering, then $\{f^{-1}(U_\alpha)\}$ is an open covering of $E$ having no finite subcovering. □

In the important space of real numbers $\mathbb{R}^1$, it is not difficult to prove that the connected sets are intervals (finite or infinite, open, half-open, or closed). Thus every connected subset of $\mathbb{R}^1$ is of one of the following nine forms: $(-\infty, \infty)$, $(-\infty, a)$, $(-\infty, a]$, $(a, \infty)$, $[a, \infty)$, $(a, b)$, $[a, b)$, $(a, b]$, $[a, b]$, where $a$ and $b$ are given real numbers.

That fact provides an important principle, known as the *intermediate value theorem*.

**Theorem 3.13.** *If $E$ is a connected set, $x$ and $y$ are any two points of $E$, $f : X \to \mathbb{R}^1$ is continuous, and $f(x) \le z \le f(y)$, then there is a point $t \in E$ such that $f(t) = z$.*

PROOF. If $z = f(x)$, take $t = x$. If $z = f(y)$, take $t = y$. In all other cases $f(x) < z < f(y)$. If there were no such point $t \in E$, the open sets $U = f^{-1}(z, \infty)$ and $V = f^{-1}(-\infty, z)$ would disconnect $E$, contrary to hypothesis. □

**Theorem 3.14.** *If $E$ is a compact set and $f : X \to \mathbb{R}^1$ is continuous, there are points $a \in E$ and $b \in E$ such that $f(a) \le f(x) \le f(b)$ for all $x \in E$.*

In other words, a continuous real-valued function on a compact set assumes maximum and minimum values.

PROOF. Since $f(E)$ is compact, it is bounded. Let $L$ and $M$ be its greatest lower bound and least upper bound. These belong to the closed set $f(E)$, since otherwise they would be in the exterior of $f(E)$, and there would be intervals $(L - \varepsilon, L + \varepsilon)$ and $(M - \varepsilon, M + \varepsilon)$ contained in $\mathbb{R}^1 \setminus f(E)$, and in that case, $L + \varepsilon/2$ would be a larger lower bound for $f(E)$ and $M - \varepsilon/2$ a smaller upper bound. Thus, there are points $a$ and $b$ such that $f(a) = L$ and $f(b) = M$, and hence $f(a) \le f(x) \le f(b)$ for all $x \in E$. □

If $X$ is a topological space, a continuous mapping $f : [a, b] \to X$, where $[a, b]$ is a closed interval of the real line, is called a *path* joining $f(a)$ and $f(b)$. If for any two points $x \in E$ and $y \in E$, there is a path in $E$ joining $x$ and $y$, $E$ is said to be *path-connected*. It is clear that a path-connected set is connected, since $f\big([a, b]\big)$ is a connected subset of $E$, and it is trivial to prove that if $F_\alpha$ is a connected set for each index $\alpha$ and any two of the sets $F_\alpha$ and $F_\beta$ have a common point, then the union of the sets $F_\alpha$ is connected.

**2.2. Shortcuts: bases and subbases.** A collection of sets $\mathfrak{B}$ is a *base* for a topology $\mathfrak{T}$ if every set in $T$ is a union of sets in $\mathfrak{B}$. For example, in the usual topology of $\mathbb{R}^n$ (or indeed, any metric space), the set of all open balls $B_r(\boldsymbol{x})$ as $r$ ranges over the positive numbers and $\boldsymbol{x}$ ranges over $\mathbb{R}^n$ is a base for the topology. A smaller base can be obtained by restricting $r$ to the positive rational numbers and $\boldsymbol{x}$ to the points all of whose coordinates are rational. For in that case, each $B_r(\boldsymbol{x})$ is a union of balls from this smaller collection, and hence every open set is also such a union. (Proof: Let $\boldsymbol{y} \in B_r(\boldsymbol{x})$, that is, $|\boldsymbol{y} - \boldsymbol{x}| < r$. Let $q$ be a positive rational number smaller than $r - |\boldsymbol{y} - \boldsymbol{x}|$. Then for each coordinate $i$, there is a point $\boldsymbol{z}$ all of whose coordinates are rational and such that $|z^i - y^i| < q/n$. We claim that $\boldsymbol{y} \in B_q(\boldsymbol{z}) \subseteq B_r(\boldsymbol{x})$. The first relation here is a trivial computation, since $|\boldsymbol{y} - \boldsymbol{z}| \leq |z^1 - y^1| + \cdots + |z^n - y^n| < q$. As for the second, if $\boldsymbol{u} \in B_q(\boldsymbol{z})$, then $|\boldsymbol{z} - \boldsymbol{x}| \leq |\boldsymbol{y} - \boldsymbol{x}| + |\boldsymbol{y} - \boldsymbol{z}| < |\boldsymbol{y} - \boldsymbol{x}| + q < |\boldsymbol{y} - \boldsymbol{x}| + r - |\boldsymbol{y} - \boldsymbol{x}| = r$.)

Thus, $\mathbb{R}^n$ actually has a countable base of open sets. Such a space is said to be *second-countable*, which naturally raises the question of what is meant by *first-countable*. A collection $\mathfrak{N}_x$ whose elements are neighborhoods of a point $x$ in a topological space $X$ is called a *neighborhood base at* $x$ if every neighborhood of $x$ contains a set belonging to $\mathfrak{N}_x$. If $\mathfrak{N}_x$ is a neighborhood base, so is the set $\mathfrak{N}_x^o$ consisting of the open sets in $\mathfrak{N}_x$, and hence we may always assume that a neighborhood base at $x$ consists of open neighborhoods of $x$. Just to be clear, we shall call it an *open neighborhood base*, since a neighborhood $N$ of $x$ may in general contain some or all of its boundary points. If $\mathfrak{N}_x$ is countable, then $X$ is *first-countable* at $x$, and we don't generally consider any spaces that are first-countable at some points but not others. Hence we say $X$ is *first-countable* if there is a countable neighborhood base at each of its points. It is trivial to prove that a second-countable space is also first-countable. Any metric space is first-countable, since a neighborhood base at $x$ can be formed from the open balls $B_q(x)$ as $q$ ranges over the positive rational numbers. Given an open neighborhood base $\mathfrak{N}_x$ at each point $x$ in a topological space $X$, the union of all these open neighborhood bases as $x$ ranges over $X$ is a base for the topology of $X$.

An arbitrary non-empty collection $\mathfrak{B}$ of subsets of $X$ will be the base of a uniquely determined topology $\mathfrak{T}$ provided the following condition is met: *If* $x \in U \cap V$*, where* $U \in \mathfrak{B}$ *and* $V \in \mathfrak{B}$*, there exists* $W \in \mathfrak{B}$ *such that* $x \in W \subseteq U \cap V$. This condition shows that the intersection $U \cap V$ is the union of sets in $\mathfrak{B}$, and obviously the union of any collection of unions of sets in $\mathfrak{B}$ is a union of sets in $\mathfrak{B}$. The only condition yet to be verified in order to prove that class of unions of sets in $\mathfrak{B}$ is a topology is that $\varnothing$ and $X$ are such unions. These two sets could simply be adjoined by *fiat* without disturbing the condition just stated. Alternatively, we could invoke a standard convention in set theory that the intersection of an empty collection of subsets of $X$ is $X$ and the union of an empty collection of subsets of $X$ is $\varnothing$.

Very often, it is impossible to describe all the open sets in a topology $\mathfrak{T}$ explicitly, to *name* them, but it is easy to describe a base and prove that it satisfies the condition that assures that it is the base of a topology. Such a base is often given by describing an open neighborhood base $\mathfrak{N}_x$ at each point $x$. (Almost equally often, $\mathfrak{N}_x$ consists of all the open neighborhoods of $x$). A collection of collections of sets $\mathfrak{N}_x$ such that $x \in N$ for each $N \in \mathfrak{N}_x$, as $x$ ranges over $X$, will be a base provided the following condition is met: *For any open sets* $U$ *and* $V$*, if* $z \in U \cap V$*, where* $U \in \mathfrak{N}_x$ *and* $V \in \mathfrak{N}_y$*, then there exists* $W \in \mathfrak{N}_z$ *such that* $W \subseteq U \cap V$. The proof

of this statement is the same as the proof of the criterion just given for a collection of subsets of $X$ to be the base of a topology.

For first-countable spaces $X$ it is possible to state continuity in terms of sequences. A sequence $\{x_n\}_{n=1}^{\infty}$ in any topological space (first-countable or not) is said to *converge* to a point $x$ if for every neighborhood $N$ of $x$ there is an index $n_0$ such that $x_n \in N$ for all $n > n_0$. The sequence is said to be *near $x$ infinitely often* if for every neighborhood $N$ of $x$ there are infinitely many indices $n$ such that $x_n \in N$.

**Theorem 3.15.** *If the topological space $X$ is first-countable and a sequence $\{x_n\}_{n=1}^{\infty}$ is near a point $x$ infinitely often, then there is a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ that converges to $x$.*

PROOF. Let $N_1, N_2, \ldots$, be a countable neighborhood base at $x$. Then there exist indices $n_1 < n_2 < \ldots$ such that $x_{n_k} \in (N_1 \cap \cdots \cap N_k)$, and hence in particular $x_{n_k} \in N_l$ if $k \geq l$. If $N$ is any neighborhood of $x$, there exists $k_0$ such that $x \in N_{k_0} \subseteq N$. It follows that $x_{n_k} \in N$ for all $k \geq k_0$, and hence $x_{n_k}$ converges to $x$. □

In first-countable spaces, it is possible to state continuity in terms of sequences.

**Theorem 3.16.** *Let $X$ be a first-countable topological space, $Y$ a topological space, and let $f$ be a function mapping $X$ into $Y$. Then $f$ is continuous at a point $x$ if and only if for every sequence $\{x_n\}_{n=1}^{\infty}$ that converges to $x$, the sequence $\{f(x_n)\}_{n=1}^{\infty}$ converges to $f(x)$.*

PROOF. Suppose $f$ is not continuous at $x$, and let $N_1, \ldots, N_n, \ldots$ be a neighborhood base at $x$. Then there is a neighborhood $N_{f(x)}$ of $f(x)$ such that $M = f^{-1}(N_{f(x)})$ is *not* a neighborhood of $x$, that is $M$ does not contain any of the sets $N_n$. Let $x_n \in (N_1 \cap \cdots \cap N_n) \setminus M$ for each $n$. Then $x_n \to x$, since any neighborhood $N$ of $x$ contains $N_1 \cap \cdots \cap N_n$ for some $n$, and hence contains $x_m$ for all $m \geq n$. But $f(x_n)$ does not converge to $f(x)$, since there is no $n$ for which $f(x_n) \in N_{f(x)}$.

Conversely, if $f$ is continuous at $x$ and $\{x_n\}_{n=1}^{\infty}$ converges to $x$, it is easy to see that $\{f(x_n)\}_{n=1}^{\infty}$ converges to $f(x)$. For if $N$ is any neighborhood of $f(x)$, then $f^{-1}(N)$ is a neighborhood of $x$, and so there exists $n_0$ such that $x_n \in f^{-1}(N)$ for all $n > n_0$. This says precisely that $f(x_n) \in N$ for all $n > n_0$. (This half of the theorem does not use the hypothesis of first-countability.) □

*Subbases.* Given an arbitrary collection $\mathfrak{A}$ of subsets of $X$, there is a unique smallest topology in which all the sets in $\mathfrak{A}$ are open sets. A pure existence proof of this fact goes as follows: There is at least one topology containing $\mathfrak{A}$, namely the discrete topology in which every subset of $X$ is an open set. Now the intersection of any collection of topologies is itself a topology. (Since each topology contains $\varnothing$ and $X$ and is closed under finite intersections and arbitrary unions, the same is true of the collection of sets common to all the topologies in the collection.) Hence the intersection of all topologies containing $\mathfrak{A}$ is a topology, and obviously the unique smallest topology with this property. The collection of sets $\mathfrak{A}$ is said to *generate* the topology, or to be a *subbase* of it.

A more "constructive" way of proceeding is just to prove that the collection of all unions of finite intersections of sets in $\mathfrak{A}$ is a topology, and that this collection must be contained in any topology that contains $\mathfrak{A}$. As before, we can get $\varnothing$ and $X$

into this collection by considering an empty union and an empty intersection. The fact that this collection is closed under arbitrary unions is immediate. The fact that it is closed under intersections follows from the distributive law in set theory: $U \cap (V \cup W) = (U \cap V) \cup (U \cap W)$.

In verifying that a function $f : X \to Y$ is continuous, it suffices to verify that $f^{-1}(U)$ is open just for the open sets $U$ in a subbase of the topology of $Y$, since inverse functions preserve set operations. If this verification has been carried out for a subbase of the topology, then every open set $U$ in $Y$ is a union of finite intersections of subbasic open sets, and hence $f^{-1}(U)$ is the corresponding union of finite intersections of the inverse images of these sets. It is therefore open.

**2.3. Category.** A set $E$ such that $\operatorname{int}(E^c) = \varnothing$ is said to be *nowhere dense*. Obviously, if $E$ is nowhere dense, so is its closure $E^c$.

**Definition 3.9.** A subset $E$ of a topological space is said to be *of first category on* $X$ if $E$ is a countable union $E = E_1 \cup \cdots \cup E_n \cup \cdots$ of nowhere-dense sets $E_n$. A set that is *not* of first category is *of second category on* $X$.

This notion is due to René Baire (1874–1932), who used it to prove that the pointwise limit of a sequence of continuous functions is continuous at "most" points. Felix Hausdorff criticized the terminology, saying it was "colorless." Whether a set $E$ is of first or second category depends very much on the space it is embedded in. For example, the set $\mathbb{Z}$ of integers is nowhere dense in the set $\mathbb{Q}$ of rational numbers and hence of first category on $\mathbb{Q}$, but $\mathbb{Z}$ is of second category on itself, since the singleton set $\{n\}$ is an open set in that space. Thus, the only nowhere-dense subset of $\mathbb{Z}$ is the empty set, and any non-empty subset of $\mathbb{Z}$ is of second category on $\mathbb{Z}$.

The space of rational numbers $\mathbb{Q}$ is of first category even on itself, since a singleton point $\{r\}$ is nowhere dense in $\mathbb{Q}$, and $\mathbb{Q}$ is the countable union of these singleton sets.

To state our main theorem, which will imply that every (finite-dimensional) manifold is of second category on itself, we need one more concept, plus a lemma.

**Definition 3.10.** A topological space $X$ is *locally compact* if each point $x$ has a compact neighborhood.

**Lemma 3.1.** *Let $X$ be a Hausdorff space, $x$ a point of $X$ having a compact neighborhood $N$, and $V$ any open set containing $x$. Then there exists an open set $U$ containing $x$ whose closure is compact and contained in $V$.*

PROOF. The open set $W = V \cap \operatorname{int}(N)$ contains $x$, and its closure $W^c$, being a closed subset of the compact set $N$, is compact. For the same reason, the closed set $W^c \setminus V$ is compact and is disjoint from the compact set $\{x\}$. It follows from what was proved above that there is a pair of disjoint open sets $V_1$ and $V_2$ such that $x \in V_1$ and $W^c \setminus V \subseteq V_2$. Let $U = V_1 \cap W$. Then $x$ belongs to the open set $U$, and $U^c$ is contained in $W^c$ and also in $V_1^c$, which is disjoint from $W^c \setminus V$. It follows that $U^c$ is contained in $W^c \cap V$ and *a fortiori* is contained in $V$. □

**Theorem 3.17.** *A locally compact Hausdorff space is of second category on itself.*

PROOF. Let $M_1, M_2, \ldots$ be nowhere-dense subsets of a locally compact Hausdorff space $X$. We intend to prove that the union of the sets $M_n$ is not all of $X$. Replacing $M_n$ by its closure $M_n^c$ if necessary, we can assume that these sets are closed, and

replacing $M_n$ by $M_1 \cup \cdots \cup M_n$ if necessary (since a finite union of nowhere-dense sets is nowhere dense), we can assume that $M_1 \subseteq M_2 \subseteq \cdots$. We note that $\text{int}\,(X) = X$, so that $X$ itself is *not* nowhere dense, and hence none of the individual sets $M_n$ contains all of $X$.

Now let $x_1$ be any point of $X$. If $x_1$ does not belong to any of the sets $M_n$, then their union is not all of $X$, and we are finished.

Assume now that $x_1 \in M_{n_1}$, and hence that $x_1$ belongs to $M_n$ for all $n \geq n_1$, Since $M_{n_1}$ is not all of $x$, there exists a point $x_2 \in X \setminus M_{n_1}$. By the lemma, there is an open set $U_2$ containing $x_2$ whose closure is compact and contained in the open set $X \setminus M_{n_1}$, so that $U_2^c \cap M_{n_1} = \varnothing$ If $x_2$ does not belong to any of the sets $M_n$, then, once again, we are finished, having exhibited a point not in any $M_n$.

If $x_2 \in M_{n_2}$ for some $n_2 > n_1$, we observe that $U_2$ is *not* contained in $M_{n_2}$, since $U_2$ is a non-empty open set and $M_{n_2}$ contains no non-empty open sets, being nowhere dense. Let $x_3 \in U_2 \setminus M_{n_2}$. Let $U_3$ be an open set containing $x_3$ whose whose closure is compact and contained in $U_2 \setminus M_{n_2}$, as above.

Continuing in this way, either at some finite stage we arrive at a point $x_k$ that is not in any $M_n$, and we are done at that point, or else we produce a shrinking sequence of nonempty compact sets $U_n^c$, each of which has non-empty interior:

$$U_2^c \supseteq U_2 \supseteq U_3^c \supseteq U_3 \cdots \supseteq U_n^c \supseteq U_n \supseteq U_{n+1}^c \supseteq \cdots$$

with the property that $U_k^c \cap M_{n_{k-1}} = \varnothing$ for each $k = 2, 3, \ldots$. We claim that the intersection

$$\bigcap_{n=2}^{\infty} U_n^c$$

is not empty.

If this intersection is empty, the complements $X \setminus U_3^c$, $X \setminus U_4^c$,..., $X \setminus U_n^c$ form an open covering of $X$ and *a fortiori* an open covering of the compact set $U_2^c$. But then some finite subcovering must exist, which, since the sets are nested, means that $U_2^c$ is contained in $X \setminus U_n^c$ for some $n$, and hence that $U_n^c$ is contained in $X \setminus U_2^c$. Since $U_n^c \subseteq U_2^c$, it then follows that $U_n^c = \varnothing$, contrary to hypothesis.

If $x$ belongs to all the sets $U_n^c$, then $x$ does not belong to any set $M_n$. Thus, in all cases, $X$ is not the union of the nowhere-dense sets $M_n$. □

For metric spaces $X$, even those that are not locally compact, there is an analog of the preceding theorem, provided the metric space has the following property: *If* $\overline{B}_{r_n}(x_n) = \{x \in X : d(x, x_n) \leq r_n\}$ *is a nested sequence of closed balls (that is* $\overline{B}_{r_{n+1}}(x_{n+1}) \subseteq \overline{B}_{r_n}(x_n)$ *for each n), and* $r_n \downarrow 0$*, then the intersection of the sets* $\overline{B}_{r_n}(x_n)$ *is non-empty.* A metric space with this property is said to be *complete*.[5] The analog of the preceding theorem, known as the *Baire category theorem*, is proved following exactly the same argument given to prove the preceding theorem, which we shall call the Baire category theorem for locally compact Hausdorff spaces.

The Baire category theorem has a number of uses, especially in proving the closed-graph theorem and the open-mapping theorem in functional analysis. It also proves the "existence" of many cautionary counterexamples in analysis. It can be shown, for example, that the set of bounded continuous real-valued functions on

[5] The nested-ball condition stated here is equivalent to the statement that every *Cauchy* sequence in $X$ converges. A sequence $\{x_n\}_{n=1}^{\infty}$ in $X$ is a Cauchy sequence if for every $\varepsilon > 0$ there exists an index $N$ such that $d(x_m, x_n) < \varepsilon$ if $m > N$ and $n > N$. Completeness says that if the points of a sequence in $X$ are "getting close to one another" as the index increases, then they must all be "getting close" to some point of $X$. Hence the name *completeness*.

the real line having a derivative at even one point is of first category in the complete metric space of such functions. Hence there exist bounded continuous real-valued functions of a real variable having no derivative at any point.

**2.4. Paracompactness.** The concept of compactness has already made the subject of topology arcane enough for most people. It is therefore with great regret that we are forced to introduce a concept that is yet more arcane. To do so, we need three definitions:

**Definition 3.11.** Let $\{U_\alpha\}$ be an open covering of a set $E$. A second open covering $\{V_\beta\}$ is a *refinement* of the first if for every $\beta$ there exists $\alpha$ such that $V_\beta \subseteq U_\alpha$.

**Definition 3.12.** An open covering $\{U_\alpha\}$ of a set $E$ is *locally finite* if each $x \in E$ has a neighborhood $N_x$ such that $N_x \cap U_\alpha = \varnothing$ for all but a finite set of indices $\alpha$.

**Definition 3.13.** A subset $E$ of a topological space $X$ is *paracompact* if every open cover of $E$ has a locally finite refinement.

We need the concept of paracompactness because a paracompact $C^\infty$-manifold $\mathfrak{M}$ has a partition of unity subordinate to any open covering of $\mathfrak{M}$. These concepts are discussed in Appendix 4. Right now, we need a simple sufficient criterion for paracompactness. To that end, we require yet one more definition.

**Definition 3.14.** A topological space $X$ is *$\sigma$-compact* if there is a sequence of compact sets $\{C_n\}_{n=1}^\infty$ whose union is $X$.

All of the spaces we shall encounter, in particular finite-dimensional manifolds, are $\sigma$-compact. An example of a space that isn't is provided by any uncountable set endowed with the discrete topology.

**Theorem 3.18.** *A locally compact, $\sigma$-compact Hausdorff space $X$ is paracompact.*

PROOF. We shall construct open sets $U_n$, $n = -1, 0, 1, 2, \ldots$, having the following three properties:

1. The closure $U_n^c$ of each set is compact.
2. $$\varnothing = U_{-1} = U_{-1}^c = U_0 = U_0^c \subseteq U_1 \subseteq U_1^c \subseteq U_2 \subseteq U_2^c \subseteq \cdots .$$
3. $$X = \bigcup_{n=1}^{\infty} U_n .$$

Once that is done, for any covering of $X$ by open sets $\{U_\alpha\}$ and each $n = 1, 2, \ldots,$, there is a finite subcollection of the sets $U_\alpha$, relabeled as $U_{n1}, \ldots, U_{nk_n}$, that covers the compact set $U_n^c \setminus U_{n-1}$. Then the sets $V_{nj} = U_{nj} \cap (U_{n+1} \setminus U_{n-2}^c)$, $j = 1, 2, \ldots, k_n$, $n = 1, 2, 3, \ldots$, provide a locally finite refinement of the original covering.

It is obvious that $V_{nj}$ is an open subset of $U_{nj}$ for each $n$ and $j$, and hence that the sets $V_{nj}$ constitute a refinement of the original covering. We need to show that this collection covers $X$ and is locally finite.

Since

$$\bar{U}_n^c \setminus U_{n-1} \subseteq U_{n+1} \setminus U_{n-2}^c$$

and also

$$U_n^c \setminus U_{n-1} \subseteq \bigcup_{j=1}^{k_n} U_{nj}\,,$$

it follows that

$$U_n^c \setminus U_{n-1} \subseteq \Big(\bigcup_{j=1}^{k_n} U_{nj}\Big) \cap \big(U_{n+1} \setminus U_{n-2}^c\big) = \bigcup_{j=1}^{k_n} V_{nj}\,,$$

and therefore that

$$X = \bigcup_{n=1}^{\infty} U_n \setminus U_{n-1} \subseteq \bigcup_{n=1}^{\infty} U_n^c \setminus U_{n-1} \subseteq \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{k_n} V_{nj}\,,$$

so that the sets $V_{nj}$ cover $X$.

It remains to be shown that the collection of open sets $V_{nj}$ is a locally finite covering.

To that end, we note that $V_{nj}$ is disjoint from $U_m$ if $m \leq n-2$. For $V_{nj}$ is disjoint from $U_{n-2}^c$, which contains $U_m$ for all $m \leq n-2$. Also, $V_{nj}$ is contained in $U_k$ for all $k > n$ and hence is disjoint from $X \setminus U_k$ for all such $k$. It follows that if $V_{nj}$ intersects the set $U_m \setminus U_{m-2}^c$, then on the one hand (because $V_{nj}$ intersects $U_m$), we have $n \leq m+1$. On the other hand, if $n \leq m-3$, then $V_{nj}$, being contained in $U_{n+1}$ is contained in $U_{m-2}$, which is contained in $U_{m-2}^c$, and hence $V_{nj}$ is disjoint from $U_m \setminus U_{m-2}^c$. It follows that $V_{nj}$ intersects the open set $U_m \setminus U_{m-2}^c$ only for $n$ satisfying $m-2 \leq n \leq m+1$, that is, only for a finite set of indices $n$ and $j$. Now every point $x \in X$ belongs to $U_m \setminus U_{m-1}$ for a unique positive integer $m$, and $U_m \setminus U_{m-1}$ is contained in the open set $U_m \setminus U_{m-2}^c$. We have now shown that the sets $V_{nj}$ are a locally finite refinement of the covering $\{U_\alpha\}$.

We will be finished if we construct the desired sets $U_n$, $n = -1, 0, 1, \ldots$. We begin with the fact that $X$ is $\sigma$-compact, so that there is a sequence of compact sets $\{C_n\}_{n=1}^{\infty}$ such that

$$X = \bigcup_{n=1}^{\infty} C_n\,.$$

Since $X$ is locally compact, each point $x \in X$ has a compact neighborhood $N_x$, and $x \in \operatorname{int}(N_x) = W_x$. For some finite set of points $x_{11}, \ldots, x_{1r_1}$, the open sets $W_{x_{1j}}$ cover $C_1$. Let

$$U_1 = \bigcup_{j=1}^{r_1} W_{x_{1j}}\,.$$

Then the set $C_2 \cup U_1^c$, being a closed subset of the compact set $C_2 \cup \big(\cup_{j=1}^{r_1} N_{x_{1j}}\big)$, is compact, and therefore covered by a finite collection of sets $W_{x_{2j}}$, $j = 1, 2, \ldots, r_2$, whose union becomes the set $U_2$ in our construction. It is now obvious how to continue the construction, and the proof is complete. □

It follows from this theorem that $\mathbb{R}^n$ is paracompact. All the manifolds we shall consider in the next appendix (and throughout this book) will be assumed to be $\sigma$-compact Hausdorff spaces. The fact that they are (finite-dimensional) manifolds guarantees that they are locally compact, and hence paracompact. But there is nothing in the definition of the concept of a manifold that requires a manifold to be a Hausdorff space or to be $\sigma$-compact. Those are independent assumptions introduced to avoid the chamber of horrors we are now going to look into briefly.

**2.5. Pathological topology. Topological pathology.** When a topological space does not have a Hausdorff topology, the topics of compactness, connectedness, convergence, and continuity can become quite trivial. In the extreme case of the indiscrete topology, for example, every sequence $\{x_n\}_{n=1}^{\infty}$ in $X$ converges simultaneously to every point $x \in X$. Every subset is compact and connected, and the only continuous mappings from $X$ to any Hausdorff space are those that assume only one value.

Another example, almost as extreme, occurs when $X$ is any infinite set and the topology $\mathfrak{T}$ consists of $\varnothing$ together with all sets whose complements have lower cardinality than $X$ itself. We have already seen two examples of this topology. When $X$ is countably infinite, the open sets are $\varnothing$ and the *co-finite* sets $E$ (sets such that $X \setminus E$ is finite. If $X$ is uncountable, the open sets may be $\varnothing$ and the *co-countable* sets $E$, for which $X \setminus E$ is countable. These conditions certainly define two topologies, but they do not have the Hausdorff property, since both of these topologies have the property that any two non-empty open sets have a non-empty intersection.

It is easy to show that in the co-finite topology every infinite subset of $X$ is connected. For if $E$ is infinite and $U$ and $V$ are open sets that both intersect $E$, the intersection $U \cap V \cap E$ cannot be non-empty, since the complement of $U \cap V$ is finite and $E$ is infinite. On the other hand, a finite set $E$ containing more than one point cannot be connected, since if $E = \{x_1, x_2, \ldots, x_n\}$, where $n \geq 2$, then $U = X \setminus \{x_1\}$ and $V = X \setminus \{x_2, \ldots, x_n\}$ are open sets such that $E \subseteq U \cup V$, $E \cap U \neq \varnothing \neq E \cap V$, and $E \cap U \cap V = \varnothing$.

In this topology, every subset of $X$ is compact, since if $E$ is covered by a collection of open sets, any one of the sets contains all but finitely many points of $E$, and then only finitely many more sets are needed to cover $E$.

Finally, in this topology, the only continuous real-valued function $f : X \to \mathbb{R}^1$ is a constant. For if $f(x) \neq f(y)$, let $U$ and $V$ be disjoint open intervals containing $f(x)$ and $f(y)$ respectively. Then $x \in f^{-1}(U)$, $y \in f^{-1}(V)$, and $f^{-1}(U) \cap f^{-1}(V) = f^{-1}(U \cap V) = \varnothing$. Thus we get two disjoint, non-empty open subsets of $X$, and that is impossible in this topology

Likewise, in the co-countable topology every uncountable subset is connected, and, as above, any countable set containing more than one point is not connected. In this topology the only compact sets are finite (these are obviously compact in any topological space whatsoever). For if $E$ is an infinite set, then $E$ contains a countable subset $\{x_1, x_2, \ldots, \}$, and then the open sets $X \setminus \{x_n, x_{n+1}, \ldots, \}$ cover $E$, but no finite subcollection will do so. And once again, as above, the only continuous real-valued functions defined on $X$ are constants.

In a general topological space, it is possible for a sequence $\{P_n\}_{n=1}^{\infty}$ to be in each neighborhood of a point $P$ infinitely often, yet have no subsequence that converges to $P$. Here is an example.

**Example 3.7.** Let the topological space $X$ consist of all subsets of the closed interval $[0, 1]$, that is, $X = 2^{[0,1]}$. A basic neighborhood of a set $A \subseteq [0, 1]$ is the set of subsets $N(A, S)$ indexed by a finite set $S \subset [0, 1]$, say $S = \{x_1, \ldots, x_r\}$, where the points $x_i$ are all distinct, and defined as

$$N(A, S) = \{B \subseteq [0, 1] : \ (A + B)S = \varnothing \}.$$

Here $A + B$ is the symmetric difference: $A + B = (A \setminus B) \cup (B \setminus A)$, otherwise described as the set of points belonging to exactly one of the two sets (exclusive

or, in the language of logic). The product is set intersection. It is well-known that $X$ forms a Boolean ring under these operations, with the empty set $\varnothing$ as identity for addition and the entire space $[0,1]$ as the identity for multiplication.[6] The statement $B \in N(A,S)$ can be phrased by saying that each element of $S$ is in either both of $A$ and $B$ or in neither of them. In other words, in order for the set $B$ to belong to the neighborhood $N(A,S)$ of $A$, it is necessary and sufficient that $B$ "agree" with $A$ about the inclusion or exclusion of each point of $S$.

If $C \in N(A,S) \cap N(B,T)$, where $S$ is as above, and $T = \{y_1, \ldots, y_s\}$, it is easy to see that $N(C, S \cup T) = N(A,S) \cap N(B,T)$, so that the sets $N(A,S)$ really do form the basis of a topology on $X$.

For each pair $(k,m)$ of positive integers with with $k \le m$, let $C_{mk}$ be the closed interval

$$C_{mk} = \Big[\frac{k-1}{m}, \frac{k}{m}\Big] .$$

For each $m$, there are $m$ sets $C_{mk}$, each being a closed interval of length $1/m$. We note that, since the endpoints of these intervals are rational numbers, for each irrational number $x \in [0,1]$ and each positive integer $m$ there exists a unique index $k(m,x)$ such that $x \in C_{mk(m,x)}$. Therefore, for any finite set of irrational numbers $x_1, \ldots, x_r$, there is a finite set of indices $k_1, \ldots, k_r$ (not necessarily distinct) such that $\{x_1, \ldots, x_r\} \subset C_{mk_1(m,x_1)} \cup \cdots \cup C_{mk_r(m,x_r)}$. Let $I_m$ be the set whose elements are the unions of the sets $C_{mk}$, so that $I_m$ contains $2^m$ elements (counting the empty union, which is the empty set).

We now define a sequence $\{P_n\}_{n=1}^{\infty}$ as follows: The sets $P_n$, $2^m - 1 \le n \le 2^{m+1} - 2$, $m = 1, 2, 3, \ldots$, are the elements of $I_m$, taken in some fixed ordering. (The specific ordering is not important.)

Now consider the set $P$ consisting of the irrational numbers in $[0,1]$.

**Theorem 3.19.** *The sequence $\{P_n\}$ is infinitely often inside every neighborhood of $P$, yet has no subsequence that converges to $P$.*

Proof. To prove the first part of this claim, consider a basic neighborhood of $P$, say $N(P, \{x_1, \ldots, x_r\})$.

Let $\delta = \min\{|x_i - x_j|, i \ne j\}$. If $m_0$ is a fixed positive integer larger than $1/\delta$, then for each $m \ge m_0$ and each irrational number $x_i$, the set $C_{mi(m,x_i)}$, which is determined uniquely by the condition that $x_i \in C_{mi(m,x_i)}$, does not contain any of the points $x_j$ with $j \ne i$. If $E_m$ is the union of the sets $C_{mi}(m, x_i)$ for which $x_i$ is irrational, then $E_m \in N\big(P, \{x_1, \ldots, x_r\}\big)$. Since $E_m$ is one of the sets $P_n$ for $2^{m-1} \le n < 2^m$, it follows that the set $P_n$ belongs to this neighborhood for infinitely many $n$.

The first part of the claim is now established, and we turn to the second part, that is, proving that no subsequence of $\{P_n\}$ can converge to $P$.

Suppose $\{P_{n_m}\}_{m=1}^{\infty}$ converges to $P$. We claim that this means $P$ is the inferior limit of this sequence of sets, that is,

$$P = \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} P_{n_k} = \bigcup_{m=1}^{\infty} B_m .$$

[6] Although we do not need this fact, the space $X$ is a commutative and associative ring under these two operations, and every element is idempotent and of order 2, that is, $AA = A$ and $A + A = \varnothing$.

To prove that claim, let $x$ be any point in $[0,1]$ and consider the neighborhood $N(P,\{x\})$. There are two cases to consider. Assume first that $x \in P$. Then, by definition of the statement that $P_{n_k} \in N(P,\{x\})$, for some index $k_0$, we have $x \in P_{n_k}$ for all $k \geq k_0$. That in turn means $x \in B_{k_0}$, and hence $x \in \bigcup_{m=1}^{\infty} B_m$.

Next assume $x \notin P$. Now for all $k \geq k_0$, again by definition of the statement $P_{n_k} \in N(P,\{x\})$ we have $x \notin P_{n_k}$. Then for even stronger reasons, for all $m$, $x \notin B_m$, and therefore $x \notin \bigcup_{m=1}^{\infty} B_m$. The equality is now proved.

Since $P_{n_k}$ is a finite union of closed intervals, the set $B_m$ is an intersection of compact sets, and hence closed. On the other hand, since the union of the sets $B_m$ is the set of irrational numbers, $B_m$ contains no intervals of positive length. (An interval of positive length contains rational numbers.) Thus, the sets $B_m$ are nowhere-dense in $[0,1]$.

Now let $r_1, \ldots, r_m, \ldots$ be an enumeration of the rational numbers in $[0,1]$, and let $E_m = B_m \cup \{r_m\}$, so that $E_m$ is a nowhere-dense closed set. It now follows that the interval $[0,1]$ is the (countable) union of the sets $E_m$. But that conclusion contradicts the Baire category theorem, since $[0,1]$ is a locally compact Hausdorff space. □

## 3. Problems

**Problem 3.1.** Prove that $\partial E = \partial(X \setminus E)$.

**Problem 3.2.** Prove that a subset $E$ of a topological space $X$ is closed if and only if $\partial E \subseteq E$.

**Problem 3.3.** Prove that $(E^c)^c = E^c$, and that $\varnothing^c = \varnothing$.

**Problem 3.4.** Prove that $(E \cup F)^c = (E^c \cup F^c)$.

**Problem 3.5.** Prove that if $F_\alpha$ is a connected set for each $\alpha \in A$ and $F_\alpha \cap F_\beta \neq \varnothing$ for one index $\beta \in A$ and every $\alpha \in A$, then the union of the sets $F_\alpha$ is connected.

**Problem 3.6.** Prove that if $E$ and $F$ are compact subsets of $X$, then $E \cup F$ is also.

**Problem 3.7.** Let $A$ be a subset of the real line and $-A = \{-x : x \in A\}$. Prove that $-A$ is bounded below (resp. above) if and only if $A$ is bounded above (resp. below) and that the greatest lower bound (resp. least upper bound) of $-A$ is $-b$, where $b$ is the least upper bound (resp. greatest lower bound) of $A$.

**Problem 3.8.** Prove that the $L_\infty$ norm $\|f\|_\infty = \max\{|f(x)| : x \in [a,b]\}$ makes the space of continuous functions on $[a,b]$ into a metric space. (Show that the triangle inequality holds.)

**Problem 3.9.** Prove that $\partial(E \cap F) \subseteq \partial E \cup \partial F$, and show that in general no stronger statement than this can be made. In particular, it is emphatically not true in general that if $E \subseteq F$, then $\partial E \subseteq \partial F$.

**Problem 3.10.** Prove that $(E \cup F)' = E' \cup F'$, and in particular, if $E \subseteq F$, then $E' \subseteq F'$.

**Problem 3.11.** The closure operator $^c$ has the following properties:

1. $E \subseteq E^c$ for all $E \subseteq X$;
2. $\varnothing^c = \varnothing$.

3. $(E \cup F)^c = E^c \cup F^c$.
4. $(E^c)^c = E^c$ for all $E \subseteq X$.

As it turns out, any operator with these properties defines a topology on $X$, in which the closed sets are those sets $E$ that are left fixed by the operator, that is, $E^c = E$. Prove this fact.

**Problem 3.12.** A good example of a closure operator is the hull-kernel operator on the space of maximal ideals of a ring. For our purposes, the integers will serve as a typical ring, that is, a structure on which addition and multiplication are defined, with the usual properties of associativity, commutativity, and distributivity. (If the ring is an *algebra*, that is, in addition to its structure as a ring, it is a vector space over some field, say the real or complex numbers, an ideal is also required to be closed under scalar multiplication.)

An *ideal* $\mathfrak{A}_1$ in a ring $\mathfrak{A}$ is a subset of $\mathfrak{A}$ that is also a ring under the addition and multiplication (and scalar multiplication, if the ring is an algebra) induced from $\mathfrak{A}$ and is such that if $x \in \mathfrak{A}_1$ and $y \in \mathfrak{A}$, then $xy \in \mathfrak{A}_1$.

Obviously, $\mathfrak{A}$ itself is an ideal, as is the subring consisting of the zero element alone. We shall consider only commutative rings, that is, those in which $xy = yx$ for all $x$ and $y$. Otherwise we should have to talk about left ideals and right ideals. We shall also assume that the rings we deal with have an identity for multiplication, which we denote 1.

In the case of the ring of integers, if $\mathfrak{A}_1$ is an ideal different from the two just mentioned, then it contains a smallest positive integer $a_1$, and we claim that $\mathfrak{A}_1$ consists of precisely the multiples of $a_1$. This is easy to prove, since if $a$ is any integer, there exist integers $q$ and $r$ such that $a = qa_1 + r$ and $0 \leq r < a_1$. (That is the division-with-remainder algorithm.) If $a \in \mathfrak{A}_1$, then $r = a - qa_1$ also belongs to $\mathfrak{A}_1$, and by definition of $a_1$, we must have $r = 0$. Thus $a = qa_1$, as asserted. We call the set of multiples of $a_1$, the *principal ideal* generated by $a_1$, and we denote it $(a_1)$. Notice that if $a_1$ and $a_2$ are positive integers such that $(a_2) \subseteq (a_1)$, then $a_2 \in (a_1)$ and hence $a_2$ is a multiple of $a_1$. It is easy to see that if $a$ and $b$ are relatively prime, then $(a) \cap (b) = (ab)$.

An ideal $\mathfrak{I}$ different from $\mathfrak{A}$ is *maximal* if the only ideal of which it is a proper subset is $\mathfrak{A}$ itself. In the case of the integers, an ideal $(p)$ is maximal if and only if $p$ is prime. For $(ab) \subseteq (a)$ and $(ab) \neq (a)$ when $b > 1$, because $a \in (a) \setminus (ab)$. And if $\mathfrak{I}$ is an ideal that properly contains a prime ideal $(p)$, then it contains an integer $q$ that is not a multiple of $p$, and so is relatively prime to $p$. But then the greatest common divisor of $p$ and $q$ is 1, and this number is of the form $ap + bq$, hence in $\mathfrak{I}$. But obviously the only ideal that contains 1 is the whole ring of integers. (This is true in general, in any ring that has an identity for multiplication.)

The *kernel* of a set of ideals of a ring $\mathfrak{A}$ is the ideal that is their intersection.

In the ring of integers, the kernel of a *finite* set of ideals $(a_1), \ldots, (a_n)$ is the ideal $(a)$, where $a$ is the least common multiple of $a_1, \ldots, a_n$. The kernel of any infinite set of distinct ideals in this ring is the set $\{0\}$.

The *hull* of a subset $E$ of $\mathfrak{A}$ is the collection of all maximal ideals that contain $E$.

The hull is not an ideal; it is a set whose elements are maximal ideals. The entire set of maximal ideas of a ring is called (naturally) the *maximal ideal space* of the ring. In the case of the integers, the maximal ideal space is in one-to-one

correspondence with the set of positive prime numbers. In this example, the hull of any set $E$ is the set of all prime divisors of all elements of $E$. The hull of the kernel of a finite set of ideals $(a_1),\ldots,(a_n)$ is the set of prime ideals generated by factors of the least common multiple of $a_1,\ldots,a_n$, that is, the set of primes that divide one or more of the numbers $a_1,\ldots,a_n$. The following proposition connects all these algebraic concepts with point-set topology.

*The operator $^c$ defined on the set of maximal ideals of a ring $\mathfrak{A}$ by specifying that $E^c = \text{hull}\big(\text{kernel}\,(E)\big)$, is a closure operator.*

PROOF. It is obvious that $E \subseteq E^c$, since each maximal ideal in a collection contains the intersection of all the maximal ideals in the collection. By convention, if $E = \varnothing$ is the empty set of maximal ideals, its kernel—the intersection of the (non-existent) ideals in $E$—is the whole ring $\mathfrak{A}$. Since there are no maximal ideals containing $\mathfrak{A}$, it follows that the hull of $\mathfrak{A}$ is empty, so that $\varnothing^c = \varnothing$.

Next, if $\mathfrak{A}_1 \in E^c$, then $\mathfrak{A}_1$ contains the intersection of all the ideals in $E$. But that means that if we adjoin $\mathfrak{A}_1$ to $E$, the intersection does not get any smaller, since it was already contained in $\mathfrak{A}_1$. Obviously, it does not get any larger either, so that the kernel of $E$ is also the kernel of $E^c$. Therefore $E^c$, the hull of the kernel of $E$, is also the hull of the kernel of $E^c$, and that means $(E^c)^c = E^c$.

Now the kernel of a union of two sets $E$ and $F$ of maximal ideals is obviously contained in the kernel of $E$ and in the kernel of $F$, and hence its hull contains the hull of the kernel of $E$ and the hull of the kernel of $F$, that is, $(E \cup F)^c \supseteq E^c \cup F^c$. When $E \subseteq F$, this relation implies $F^c \supseteq E^c \cup F^c$, so that $E^c \subseteq F^c$.

Conversely, if $\mathfrak{I}$ is a maximal ideal and $\mathfrak{I} \notin E^c \cup F^c$, then $\mathfrak{I}$ does not contain the kernel of $E$ and it does not contain the kernel of $F$. That is, there is a point $x \in \ker(E) \setminus \mathfrak{I}$ and a point $y \in \ker(F) \setminus \mathfrak{I}$. Since $\ker(E)$ and $\ker(F)$ are ideals, the point $z = yx$ belongs to $\ker(E) \cap \ker(F) = \ker(E \cup F)$. Then, either $\mathfrak{I} \notin (E \cup F)^c$, or $z \in \mathfrak{I}$. The former is what we would like to show, since we want to show that $(E \cup F)^c = E^c \cup F^c$. So, suppose $z \in \mathfrak{I}$. Now $\mathfrak{I}$ is a *maximal* ideal that does not contain $x$. That means that the smallest ideal containing $\mathfrak{I}$ and having $x$ as one of its points is the whole ring. This ideal can be described as the set of all points $w + xu$, as $w$ ranges over $\mathfrak{I}$ and $u$ ranges over the whole ring. In particular $1 - xu \in \mathfrak{I}$ for some $u$, and therefore $y - yxu \in \mathfrak{I}$. But since $yxu = zu \in \mathfrak{I}$, this means $y \in \mathfrak{I}$, contradicting the choice of $y$. Thus, we do not have $z \in \mathfrak{I}$, and we are done. □

What are the closed sets in the maximal ideal space of the integers? Describe the topology generated by this closure operator.

**Problem 3.13.** Here is an amusing example of the use of a topology to prove a seemingly unrelated fact. It is due to Hillel Furstenberg (see his article "On the infinitude of primes" in the 1955 *American Mathematical Monthly*, **62**, p. 353). Let $X = \mathbb{Z}$ be the set of integers (positive, negative, and zero). For each nonnegative integer $a$ and positive integer $d > a$, define the set

$$B(a,d) = \{a + nd : n \in \mathbb{Z}\}.$$

In other words, $B(a,d)$ is an arithmetic sequence with difference $d$; putting the matter yet another way, it is the set of all integers that are equal to $a$ modulo $d$. It is clear that if $B(a_1,d_1) \cap B(a_2,d_2)$ is non-empty, and $b$ is the smallest non-negative

integer it contains, then

$$B(a_1, d_1) \cap B(a_2, d_2) = B(b, d)\,,$$

where $d$ is the least common multiple of $d_1$ and $d_2$. This is the one condition that needs to be met in order for the sets $B(a, d)$ to form a base for a topology $\mathfrak{T}$ on $\mathbb{Z}$.

Now, we make two observations about this topology:

1. The basic open sets are both open and closed, since
$$X \setminus B(a, d) = \bigcup_{0 \le b < d\,,\, b \neq a} B(b, d)\,.$$
2. A non-empty open set in this topology is infinite (since it contains an infinite set $B(a, d)$ for some $a$ and $d$).

That being given, consider the union of all prime ideals:

$$P = \bigcup_{p \text{ prime}} B(0, p)\,.$$

Show that if the set of prime numbers were finite, then $\mathbb{Z} \setminus P$ would be a finite, non-empty open set in this topology. Hence, *there are infinitely many prime numbers!* (Just to be clear: The number 1 is *not* a prime number.)

**Problem 3.14.** A point $x$ in a topological space $X$ is *isolated* if the singleton set $\{x\}$ is an open set. Prove that a subset $E$ of $X$ that contains an isolated point $x$ of $X$ is of second category on $X$.

**Problem 3.15.** A topological space $X$ is *separable* if there is a (finite or) countable set $E = \{x_1, \ldots, x_n, \ldots,\}$ such that $E^c = X$. Such a set $E$ is said to be *dense in* $X$. Prove that a second-countable topological space is separable and first-countable.

**Problem 3.16.** Exhibit a countable dense subset of $\mathbb{R}^n$.

**Problem 3.17.** Show that a compact Hausdorff space $X$ is rich in continuous real-valued functions, in the sense that if $A$ and $B$ are any two disjoint closed subsets of $X$, there exists a continuous function $f : X \to [0, 1]$ such that $f(A) = \{0\}$ and $f(B) = \{1\}$. (*Outline of the proof:* First show that if $C_0$ is a closed subset, $U_1$ an open subset, and $C_0 \subseteq U_1$, then there exists an open subset $U_{1/2}$ such that $C_0 \subseteq U_{1/2} \subseteq U_{1/2}^c \subseteq U_1$. Then start with $C_0 = A$, $U_1 = X \setminus B$, and define by induction on $n$ sets $U_{k/2^n}$ for $0 < k < 2^n$ such that $U_{(k-1)/2^n}^c \subseteq U_{k/2^n}$. For example, if $k$ is even, then $U_{k/2^n} = U_{l/2^{n-1}}$, so that $U_{k/2^n}$ is already defined at the previous stage as an open set containing the closed set $U_{(l-1)/2^{n-1}}^c$, where $l = k/2$. And similarly if $k$ is odd, then $U_{(k-1)/2^n}^c = U_{l/2^{n-1}}^c$, so that $U_{(k-1)/2^n}^c$ is already defined at the previous stage as a closed set contained in the open set $U_{(l+1)/2^n}$. In either case, you can apply the construction just shown to get the required sequence. Then for each $x \in X$, let $f(x) = \inf\{r : x \in U_r\}$, where $r$ ranges over the binary rationals in $[0, 1]$. To show that $f$ is continuous, show that the inverse image of an interval $[0, b)$ is open, and that the inverse image of $(a, 1]$ is open. Since these sets and their finite intersections form a base of the topology of $[0, 1]$ it follows that $f$ is continuous.)

**Problem 3.18.** Let $X$ be a compact space and $Y$ a Hausdorff space, and let $f : X \to Y$ be a one-to-one continuous mapping of $X$ onto $Y$. Prove that the inverse mapping $f^{-1} : Y \to X$ is continuous. It follows easily that $X$ is also a Hausdorff space, $Y$ is a compact space, and a subset $V$ of $Y$ is open if and only

if $U = f^{-1}(V)$ is an open subset of $X$. The mapping $f$ (or $f^{-1}$) is called a *homeomorphism*. It follows that, given a compact Hausdorff topology on any set $X$, the space $X$ is not a Hausdorff space in any strictly weaker topology and not a compact space in any strictly stronger one. There can be two different compact Hausdorff topologies on a given set, but neither can be stronger than the other; that is, each contains some open sets not in the other. Compact (more generally, locally compact) Hausdorff spaces satisfy the "Goldilocks principle" when choosing topologies. (But it is not always possible to get one that meets our needs; infinite-dimensional real or complex vector spaces, for example, are not locally compact.)

**Problem 3.19.** The space $\ell^1$ is defined as the set of all sequences $A = \{a_n\}_{n=0}^{\infty}$ of complex numbers (note that the index begins with 0) such that

$$\sum_{n=0}^{\infty} |a_n| < \infty$$

with addition and scalar multiplication defined termwise and multiplication defined as *convolution*:

$$A * B = C\,,$$

where, with obvious notation,

$$c_n = \sum_{j=0}^{n} a_j b_{n-j}\,.$$

It is not difficult to show that $C \in \ell^1$ if $A \in \ell^1$ and $B \in \ell^1$. With each $A \in \ell^1$ we can associate a function $A(z)$ that is continuous in the closed unit disk $|z| \le 1$ and analytic in the open disk where $|z| < 1$ by the rule

$$A(z) = \sum_{n=0}^{\infty} a_n z^n\,.$$

We observe that $A$ can be recovered from $\widehat{A}(z)$, since $a_n = \widehat{A}^{(n)}(0)/n!$. Thus the mapping $A \mapsto \widehat{A}(z)$ is one-to-one. This rule therefore defines an isomorphism from $\ell^1$ *into* (not *onto*) the space of continuous functions in the closed disk that are analytic in its interior. (Not every continuous function of period $2\pi$ has an absolutely convergent Fourier series, even if all its Fourier coefficients of negative index are equal to zero. Thus, the mapping is not "onto.") Finally, we remark that the linear mapping $A \mapsto \widehat{A}(z)$ is continuous in the sense that $\|\widehat{A}\|_\infty = \sup|\widehat{A}(z)| \le \sum |a_n| = \|A\|_1$, where $\|A\|_1$ is defined as follows: If $A = (a_0, a_1, \dots, a_n, \dots)$, then

$$\|A\|_1 = \sum_{n=0}^{\infty} |a_n|\,.$$

It is easy to show that $\ell^1$ is a complete metric space. For, let $\{C_k\}_{k=1}^{\infty}$ be a Cauchy sequence in this metric, where $C_k = (c_{k0}, c_{k1}, \dots, c_{kn}, \dots)$. This means that for every $\varepsilon > 0$ there is an integer $N$ such that $\|C_k - C_l\|_1 < \varepsilon$ if $k > N$ and $l > N$. Then for each $n$, it follows that $|c_{kn} - c_{ln}| < \varepsilon$ also, that is, $\{c_{kn}\}_{k=1}^{\infty}$ is a Cauchy sequence of complex numbers. Since the complex numbers are a complete metric space, there is a complex number $c_n$ such that $c_{kn} \to c_n$ as $k \to \infty$. If $C = (c_0, c_1, \dots, c_n, \dots)$, it is then easy to show that $\|C_k - C\|_1$ tends to 0, that is,

$C_k \to C$. Thus, $\ell^1$ is complete. Let the sequence $\{C_k\}_{k=0}^{\infty} = \big\{\{c_{kn}\}_{n=0}^{\infty}\big\}_{k=0}^{\infty}$ (*not* a Cauchy sequence!), be given by

$$c_{kn} = \begin{cases} 1 & \text{if } n = k\,, \\ 0 & \text{if } n \neq k\,. \end{cases}$$

Show that the set $\{C_0, C_1, \dots, C_k, \dots\}$ is linearly independent and that every element $C \in \ell^1$ can be approximated with arbitrary precision by some finite linear combination of these elements. Thus they are a *geometric* basis of $\ell^1$, and its geometric dimension is therefore countably infinite.

Now let $\{C_\alpha\}_{\alpha \in A}$ be a maximal linearly independent set, that is, one for which every element $C \in \ell^1$ is *equal* to some finite linear combination of these elements: $C = t_1 C_{\alpha_1} + \cdots + t_r C_{\alpha_r}$. This set is an *algebraic* basis, and the cardinality of $A$ is the algebraic dimension of the space. Show that the cardinality of the index set $A$ is that of the continuum. *Hint:* Assume that this cardinality is a countable infinity, that is, the maximal linearly independent set just named is $\{C_{\alpha_1}, \dots, C_{\alpha_n}, \dots\}$. Show that the finite-dimensional subspace spanned by the first $n$ of these elements is nowhere-dense in $\ell^1$. Then invoke the Baire category theorem, since $\ell^1$ is the union of these subspaces.

**Problem 3.20.** As mentioned in the preceding problem, if $C$ is any element of $\ell^1$ and $z$ any complex number such that $|z| \leq 1$, the series

$$\sum_{n=0}^{\infty} c_n z^n = c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n + \cdots$$

converges absolutely and uniformly over all $z$ satisfying the stated inequality (that is, $z$ lying in the closed disk of radius 1 about 0 in the complex plane). Thus, this series defines a function $\widehat{C}(z)$ in that disk that is continuous in the closed disk and analytic—equal to the sum of its Maclaurin series—in the open disk. As a function of $C$, it turns out to be analogous to the Laplace transform, which is defined (assuming suitable convergence of the integral) as

$$L(\varphi)(z) = \int_0^{\infty} \varphi(t) e^{-tz}\, dt\,.$$

The analogy pairs $\sum$ with $\int$, the sequence $C$ with the function $\varphi$, and $z^n$ with $e^{-tz}\, dt$.

Under the convolution operation, the Banach algebra (complete normed algebra) $\ell^1$ has a multiplicative identity, namely the sequence $I = (1, 0, 0, \dots)$, and $\widehat{I}(z) \equiv 1$. Formally, if $c_0 \neq 0$, one can solve the infinite system of equations that expresses the equality $C * D = I$ as follows:

$$\begin{aligned} c_0 d_0 &= 1\,, \\ c_0 d_1 + c_1 d_0 &= 0\,, \\ c_0 d_2 + c_1 d_1 + c_2 d_0 &= 0\,, \\ &\vdots \quad \vdots \quad \vdots\,, \\ c_0 d_n + c_1 d_{n-1} + \cdots + c_{n-1} d_1 + c_n d_0 &= 0\,, \end{aligned}$$

getting a sequence $D = (d_0, d_1, \dots, d_n, \dots)$ given by

$$\begin{aligned} d_0 &= \frac{1}{c_0}\,, \\ d_1 &= -\frac{c_1 d_0}{c_0} = -\frac{c_1}{c_0^2}\,, \\ d_2 &= -\frac{c_2 d_0 + c_1 d_1}{c_0} = -\Big(\frac{c_2}{c_0^2} + \frac{c_1^2}{c_0^3}\Big),, \\ d_3 &= \frac{c_3 d_0 + c_2 d_1 + c_1 d_2}{c_0} = -\Big(\frac{c_3}{c_0^2} + \frac{c_2 c_1 + c_1 c_2}{c_0^3} + \frac{c_1^3}{c_0^4}\Big)\,, \end{aligned}$$

and so on. It is an interesting question whether $D$ belongs to $\ell^1$. Show that in general it does not by considering $C = (1, -2, 0, 0, \dots)$.

**Problem 3.21.** For each fixed $z_0$ in the closed disk, the set of all $C \in \ell^1$ such that $\widehat{C}(z_0) = 0$ is a maximal ideal $\mathfrak{I}_{z_0}$. That it is an ideal is obvious. That it is maximal follows from the expression $C = \widehat{C}(z_0)(1, 0, 0, \dots) + \big(c_0 - \widehat{C}(z_0), c_1, c_2, \dots, c_n, \dots\big) = cI + C_1$, where $c$ is the scalar constant $\widehat{C}(z_0)$ and $C_1$ is such that $\widehat{C}_1(z_0) = 0$, that is, $C_1 \in \mathfrak{I}_{z_0}$. It follows that the ideal $\mathfrak{I}_{z_0}$ has co-dimension 1 and so is maximal.

Thus, we have achieved a faithful representation of the algebra $\ell^1$ as an algebra of functions on (at least part of) its maximal ideal space. (We have not shown that every maximal ideal in $\ell^1$ is $\mathfrak{I}_{z_0}$ for some $z_0$, nor that the hull-kernel topology on the maximal ideal space is the ordinary metric topology of the closed unit disk, although both of these statements are true.)

Show that if $\widehat{C}(z)$ has no zeros in the closed disk, then the sequence $D$ defined in the preceding problem generates the analytic function $\widehat{D}(z) = 1/\widehat{C}(z)$ in the open disk, and that this function is continuous on the boundary.

Show that if $d_n \geq 0$ for all large $n$, and $\widehat{C}(z)$ has no zeros in the closed disk, then $D \in \ell^1$.

This result is a special case of the *Wiener Tauberian theorem*, named after the American mathematician Norbert Wiener (1894–1964).

## APPENDIX 4

# Manifolds

*Twist and turn as you may, you cannot avoid using analytic expressions such as, for example, power series.*

Karl Weierstrass (1815–1897). Quoted by Reinhard Siegmund-Schultze, *Ausgewählte Kapitel aus der Funkionentheorie*, Teubner, Leipzig, 1988, p. 253. My translation.

To fill the gap left in our understanding of relativistic gravitational theory, we need to introduce manifolds and tensor analysis, both of which were mentioned in Chapters 4 and 5, but neither of which was formally defined. The formal discussion of tensors will be found in Appendix 6; the present appendix is devoted to manifolds. The basic idea of a manifold was introduced by Riemann in his 1854 inaugural lecture at the University of Göttingen. There he introduced the idea of an *n-fold extended quantity*, which he called a manifold (*Mannigfaltigkeit*). The basic idea is a systematization of the idea of using parameters in geometry, which, as we saw in Chapter 5, was the basic tool used by Euler and Gauss in their brilliant work on the geometry of surfaces in $\mathbb{R}^3$. The need for such a systematization comes from three inconveniences that limited the usefulness of parameters. To list these in increasing order of seriousness, they are the following:

1. The parameters come from a space whose dimension determines the geometric dimension of the object they parameterize, but the space into which they map, the "ambient space," is a matter of arbitrary choice: Since $\mathbb{R}^m \subseteq \mathbb{R}^n$ when $n \geq m$, one must wonder whether the ambient space has any effect on the properties of the manifold. In other words, quantities having no necessary relation to the object being studied seem to be an essential part of the study.
2. There are usually places on the geometric object where the parametrization "breaks down," that is, either points corresponding to more than one set of parameter values or to no parameter values at all, or points where the mapping ceases to be differentiable. For example, when latitude and longitude coordinates are used on a sphere, the two poles have no assignable longitude. When stereographic projection is used to parameterize the sphere, the pole from which the projection is made has no coordinates.
3. It is necessary to distinguish properties that are intrinsic to the object being studied from properties of a particular parametrization of it. In physical terms, two observers using different measuring systems to study a physical object, need to decide when they are observing the same phenomenon, and whether their observations agree.

All three of these problems can be handled through the notion of a manifold, which is intuitively so like a parametrization that one's imagination really requires no adjustment to make the transition. To take the first problem as an example, as we shall see, it is really impossible to define any *specific* manifold without *some* Euclidean space playing the role of the ambient space (the Whitney Embedding theorem, a weak form of which is proved below). But the game is to avoid mentioning the ambient space and any particular parametrization when proving things about a manifold. The whole concept was created as a way of building a structure in that Platonic heaven mentioned in Appendix 3. Things that were previously defined in terms of a particular parametrization and then perhaps shown to have a meaning independent of the parametrization could now be given an "intrinsic" meaning. In a stroke of genius that resembles Gauss' magnificent achievement in defining curvature purely in terms of the metric coefficients and their partial derivatives, with no need to start from a tangent plane (see Chapter 5), mathematicians have managed to define the tangent space to a manifold without any need to use properties of whatever ambient space the manifold may be in.

On the other hand, the points where a particular parametrization breaks down normally form a very small portion of a manifold, one that can be ignored for many purposes such as integration. And, in fact, when we want to define any specific manifold, we are simply forced to resort to some parametrization. That is the meaning of the quotation from Weierstrass given above.[1] In any case, by changing to another set of parameters near a point where a parametrization ceases to be one-to-one or differentiable or both, one can get a manifold covered by patches, each of which is "diffeomorphic" to an open set in a Euclidean space of the same dimension as the manifold. That is to say, there is a one-to-one mapping of an open neighborhood of each point onto an open set in a Euclidean space. This adjustment removes the need to worry about singularities in the parametrization, even though, in informal arguments, those singularities are usually ignored. It comes at the price of some complication, of course, since on patches where two such mappings $\varphi$ and $\psi$ are defined, it is necessary to verify that the mapping $\varphi \circ \psi^{-1}$ is an infinitely differentiable mapping from one open set in Euclidean space to another.

Finally, when two different parametrizations are used, say for example $(s,t) \mapsto \boldsymbol{r}(s,t) \in \mathbb{R}^n$ and $(u,v) \mapsto \boldsymbol{w}(u,v) \in \mathbb{R}^n$, there is an obvious correspondence $(s,t) \leftrightarrows (u,v)$ established by the relation $\boldsymbol{r}(s,t) = \boldsymbol{w}(u,v)$. In this way, the pair $(s,t)$ can be regarded as functions of $(u,v)$ and vice versa. Since differential geometry is essentially all about using tangent vectors to reduce questions about the manifold to the infinitesimal level, nearly all of the most important concepts that arise involve differentiation, that is, they are defined in terms of tangent vectors and linear operators on tangent vector spaces. In that context, one can take two quantities defined in different parametrizations as being the same, if the coordinates of one can

[1] Weierstrass was speaking in the context of analytic function theory, where Cauchy had taken an abstract definition of analytic functions and then proved that they could be represented by power series. Weierstrass was pointing out that the theorem about power series representations was not really relevant to any *specific* function that one might be hoping to apply to study nature, since it wasn't possible to define such a function without either a specific formula or at least a differential equation that it had to satisfy. He thought, therefore, that one might as well save all the trouble and just start with power series representation as the defining criterion for analyticity.

be obtained from those of the other through multiplication by a Jacobian matrix

$$\begin{pmatrix} \frac{\partial u}{\partial s} & \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial s} & \frac{\partial v}{\partial t} \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{pmatrix}.$$

## 1. From Parametrizations to Manifolds

Our notation is as follows: The real $n$-dimensional Euclidean space $\mathbb{R}^n$ consists of $n$-tuples of real numbers, denoted by boldface letters. Thus we write $\boldsymbol{x} = (x^1, \ldots, x^n)$, $\boldsymbol{y} = (y^1, \ldots, y^n)$, using superscripts rather than the usual subscripts to distinguish the components of a vector. The inner (dot) product on $\mathbb{R}^n$ is denoted $\boldsymbol{x} \cdot \boldsymbol{y}$ and defined by the equation $\boldsymbol{x} \cdot \boldsymbol{y} = x^1 y^1 + \cdots + x^n y^n$, and the length of a vector is $|\boldsymbol{x}| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}$.

**1.1. General manifolds.** An *$n$-dimensional manifold* is a topological space $M$ that is "locally Euclidean" in the sense that for each point $P \in M$, there is an open neighborhood $U$ of $P$ and a homeomorphism[2] $\psi$ (called a *chart at $x$*) that maps $U$ in a one-to-one manner *onto* an open subset $V$ of $\mathbb{R}^n$. The homeomorphism allows the open set $U$ to be studied as if it were a subset of Euclidean space, and hence the informal characterization of manifolds as "locally Euclidean" spaces. Since a chart $\psi$ is a homeomorphism, it is one-to-one, and therefore the inverse function $\psi^{-1}$ is a well-defined continuous function whose domain is $V$ and whose range is $U$. Notice that for any two charts $\psi_1$ and $\psi_2$, the mapping $\psi_2 \circ \psi_1^{-1}$ is a homeomorphism whose domain is $\psi_1(U_2)$ and whose range is $\psi_2(U_1)$. Now these sets may very likely be empty, so that this "function" has an empty domain and so doesn't really map anything anywhere. In that case, we shall agree to say, as in Appendix 3, that $\psi_2 \circ \psi_1^{-1}$ is a *non-existent function*, and that any assertion made about any function whatsoever is automatically true of all non-existent functions. That will save some writing, and we won't have to be constantly saying, "if the function $\psi_2 \circ \psi_1^{-1}$ is defined." In particular, it is true to say that $\psi_1(U_2)$ is the domain of $\psi_2 \circ \psi_1^{-1}$, even if this set is empty. By the definition of a manifold $M$, there exists an open covering of $M$ by the domains of the charts $\psi$. Once we have a basic set of charts whose domains cover $M$, we naturally say that these charts form a base of the manifold structure on $M$. We then proceed to adjoin all homeomorphisms of open sets $U$ into $\mathbb{R}^n$ to the base, getting a maximal set of charts. In practice, we generally need only a base for the structure, but in theoretical reasoning, we often need to invoke a particular chart having properties that we desire, and it would be tedious to be forced to say continually, "And let this chart be added to the manifold structure."

If a manifold is not connected, it consists of disjoint pieces, like isolated universes, having no interaction or unity at all. For that reason, it is customary to require a manifold to be connected, and in that case, it can easily be proved that the dimension $n$ is necessarily the same for every chart. In our presentation, we start by requiring a single Euclidean space $\mathbb{R}^n$ to provide the range of all charts, and thus it becomes irrelevant to require the manifold to be connected. In fact, in the case of manifolds of dimension 0, the only connected one is the trivial one consisting of a single point. A general manifold of dimension 0 is simply a collection of points with the discrete topology. Only one of these is of the slightest use, and even it is rather trivial. That one is the zero-sphere $\mathbb{S}^0(r)$, consisting of the two points $\pm r$ on the real line. But differentiation makes no sense in dimension 0,

[2] See Problem 3.18 of Appendix 3 for the definition of a homeomorphism.

and we shall not use even that manifold. All of our manifolds will have positive dimension.

Manifolds are automatically locally compact spaces, since $\mathbb{R}^n$ is locally compact. By imposing the additional conditions that a manifold $M$ be a $\sigma$-compact Hausdorff space, we assure that it is also paracompact, as proved in Appendix 3. That property is important in the construction of a partition of unity, which we shall give below.

**1.2. Differentiable structures.** Not much can be done with a generic manifold, which has only the properties just given. To get any interesting results, it is necessary to assume a *differentiable structure*. A $C^r$ *manifold* is a manifold for which the mappings $\psi_2 \circ \psi_1^{-1}$ all have continuous partial derivatives of every order not greater than $r$ at every point of the open set $\psi_1(U_2)$. (Again, in verifying that this condition holds, we can ignore pairs of charts for which $\psi_2 \circ \psi_1^{-1}$ is a non-existent function.) The generic manifold just defined is referred to as a $C^0$ manifold. We allow $r = \infty$ in this case, and indeed, that is the case we are primarily interested in. That is, we want $\psi_2 \circ \psi_1^{-1}$ to have continuous partial derivatives of all finite orders. There is also a subclass of $C^\infty$ manifolds called *real-analytic* manifolds, or $C^\omega$ manifolds, for which all the functions $\psi_2 \circ \psi_1^{-1}$ are analytic at every point of their domains. The reader may be familiar with the fact that a $C^\infty$ function is not necessarily analytic. The typical example of such a function is the function of a real variable defined by

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}}, & \text{if } x \neq 0\,, \\ 0\,, & \text{if } x = 0\,. \end{cases}$$

It is easy to show by L'Hospital's rule and mathematical induction that $f^{(n)}(0) = 0$ for all $n$. If this function were analytic—that is, the sum of its Maclaurin series—it would be identically zero. But in fact it is non-zero for all $x \neq 0$.

Thus, we have a decreasing set of manifolds indexed by the extended non-negative ordinal numbers $[0, 1, 2, \ldots, \infty, \omega]$. In what follows, we shall work exclusively with the $C^\infty$ case. The reason for preferring this case to the case of $C^\omega$ manifolds, which are a more natural context in the study of Lie groups, is that we shall need the partitions of unity just mentioned, and these do not exist on non-trivial analytic manifolds. They do, however, exist on $C^\infty$ manifolds. The reason for ignoring manifolds with $r < \infty$ will be explained in the next subsection. Right now, we need to examine a few examples.

**Example 4.1.** The most trivial example of a manifold is Euclidean space $\mathbb{R}^n$ itself, covered by the domain $U = \mathbb{R}^n$ of a single basic chart $\psi(\boldsymbol{x}) = \boldsymbol{x}$.

**Example 4.2.** Possibly the most important $n$-dimensional manifold is the $n$-sphere $\mathbb{S}^n(r)$ of radius $r$ (we assume $r > 0$) in $\mathbb{R}^{n+1}$, defined as follows:

$$\mathbb{S}^n(r) = \{\boldsymbol{\xi} \in \mathbb{R}^{n+1} : (\xi^1)^2 + \cdots + (\xi^{n+1})^2 = r^2\}\,.$$

It is not possible to cover $\mathbb{S}^n(r)$ with a single basic chart, since it is a compact space, and a non-empty open subset of $\mathbb{R}^n$ is not compact. It is possible, however, to get a base consisting of the two stereographic projections from the north and south poles, that is, $\psi_N : \mathbb{S}^n(r) \setminus (0, 0, \ldots, 0, r) \to \mathbb{R}^n$ and $\psi_S : \mathbb{S}^n(r) \setminus (0, 0, \ldots, 0, -r) \to$

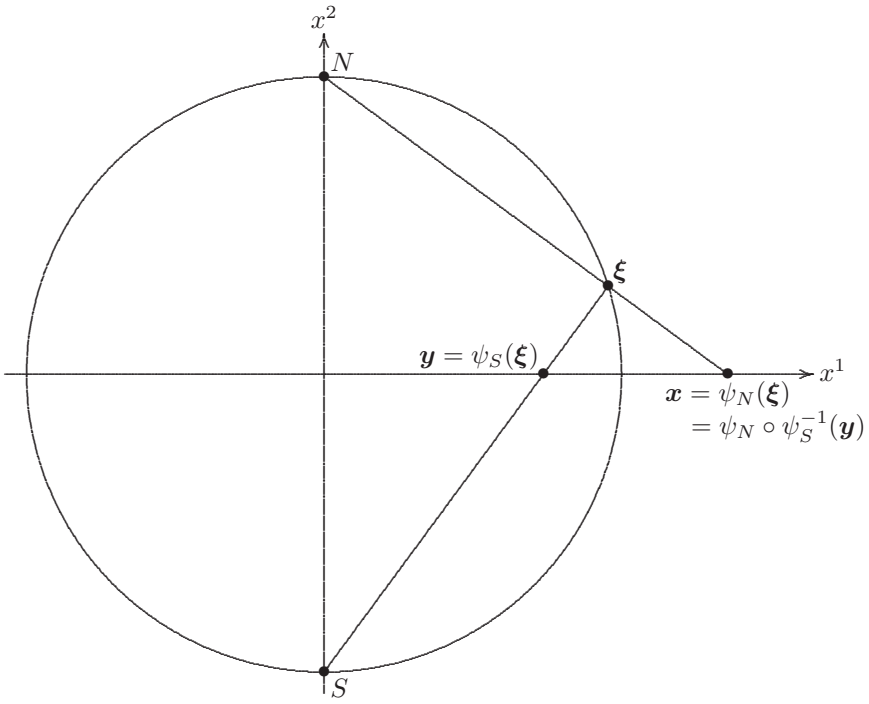


FIGURE 1. The two stereographic projections of the $n$-sphere onto Euclidean $n$-space.

$\mathbb{R}^n$, defined as follows:

$$\begin{aligned}
\psi_N(\boldsymbol{\xi}) &= \left(\frac{r\xi^1}{r-\xi^{n+1}}, \ldots, \frac{r\xi^n}{r-\xi^{n+1}}\right) = (x^1, \ldots, x^n)\,; \\
\psi_N^{-1}(\boldsymbol{x}) &= \left(\frac{2r^2x^1}{r^2+|\boldsymbol{x}|^2}, \ldots, \frac{2r^2x^n}{r^2+|\boldsymbol{x}|^2}, r\frac{|\boldsymbol{x}|^2-r^2}{|\boldsymbol{x}|^2+r^2}\right); \\
\psi_S(\boldsymbol{\xi}) &= \left(\frac{r\xi^1}{r+\xi^{n+1}}, \ldots, \frac{r\xi^n}{r+\xi^{n+1}}\right) = (x^1, \ldots, x^n)\,; \\
\psi_S^{-1}(\boldsymbol{x}) &= \left(\frac{2r^2x^1}{r^2+|\boldsymbol{x}|^2}, \ldots, \frac{2r^2x^n}{r^2+|\boldsymbol{x}|^2}, r\frac{r^2-|\boldsymbol{x}|^2}{r^2+|\boldsymbol{x}|^2}\right).
\end{aligned}$$

The composite mapping $\psi_N \circ \psi_S^{-1}$, which is the same as $\psi_S \circ \psi_N^{-1}$ has domain and range equal to $\mathbb{R}^n \setminus \{\mathbf{0}\}$ and is given by

$$\psi_N \circ \psi_S^{-1}(\boldsymbol{x}) = \frac{r^2}{|\boldsymbol{x}|^2}\boldsymbol{x}\,.$$

This mapping is analytic at each point of its domain. Thus, $\mathbb{S}^n(r)$ is actually a $C^\omega$ manifold. We shall not make any use of its analyticity, however. The two charts are illustrated for the case $n = 1$ (the circle of radius $r$) in Fig. 1.

Another possible set of basic charts on $\mathbb{S}^n(r)$ is provided by the $2n+2$ mappings $\psi_{kj}(\boldsymbol{\xi})$, $j = 1, 2$, $k = 1, \ldots, n+1$, with domains $U_{k1} = \{\boldsymbol{\xi} : \xi^k > 0\}$, $U_{k2} = \{\xi :$

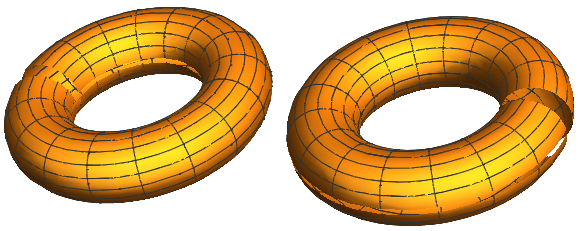

FIGURE 2. The torus cut open in two ways.

$\xi^k < 0\}$, $V_{kj} = \{\boldsymbol{x} : |\boldsymbol{x}| < r\}$, and

$$\begin{aligned} \psi_{1j}(\boldsymbol{\xi}) &= (\xi^2, \ldots, \xi^{n+1}) \\ \psi_{n+1\,j}(\boldsymbol{\xi}) &= (\xi^1, \ldots, \xi^n) \\ \psi_{kj}(\boldsymbol{\xi}) &= (\xi^1, \ldots, \xi^{k-1}, \xi^{k+1}, \ldots, \xi^{n+1}) \text{ if } 1 < k \leq n\,. \end{aligned}$$

These mappings are all homeomorphisms, since, for example, if $1 < k \leq n$ we have

$$\psi_{k1}^{-1}(\boldsymbol{x}) = (x^1, \ldots, x^{k-1}, \sqrt{r^2 - |\boldsymbol{x}|^2}, x^{k+1}, \ldots, x^n)$$

and

$$\psi_{k2}^{-1}(\boldsymbol{x}) = (x^1, \ldots, x^{k-1}, -\sqrt{r^2 - |\boldsymbol{x}|^2}, x^{k+1}, \ldots, x^n)\,.$$

It is easily verified that all of the mappings $\psi_{ki} \circ \psi_{lj}^{-1}$ are analytic, as are all the mappings $\psi_N \circ \psi_{ki}^{-1}$ and $\psi_{ki} \circ \psi_N^{-1}$. Hence, we could adjoin the charts from either set to the other set and still have the same (analytic) differentiable structure. That is why textbooks of differential geometry always assume that all possible $C^\infty$ compatible charts have already been adjoined. That way, the entire manifold structure—the underlying space and all possible charts—is a uniquely determined object.

**Example 4.3.** Another important $n$-dimensional analytic manifold is the *$n$-dimensional torus* $\mathbb{T}^n$, which is the topological product of $n$ circles. (For $n = 1$, the torus and the sphere are the same, each being an ordinary circle in the plane.) We shall confine ourselves to the case $n = 2$ in this example, which is the torus most people are familiar with. Like the $n$-sphere, $\mathbb{T}^2$ is covered, except for two points, by the domains of two charts $\psi_1$ and $\psi_2$ obtained in both cases by removing one circle of longitude (constant $\theta$ in cylindrical coordinates) and one circle of latitude (constant $r$ and $z$ in cylindrical coordinates). Provided the two circles of longitude and the two circles of latitude are different, their domains will cover all of $\mathbb{T}^2$ except the two points where the omitted circle of latitude in one chart intersects the omitted circle of longitude in the other. (It is of course trivial to introduce two more charts to cover these two points, but we are not going to bother to do so.) When one circle of longitude and one circle of latitude are removed, what remains can be pictured as the torus cut open along these lines, and it is then easy to see that this remaining portion can be "flattened out" onto a rectangle. The cut-open tori are shown in Fig. 2.

The torus has the equation $(r - a)^2 + z^2 = b^2$ in cylindrical coordinates, where $a > b$. For convenience, we'll take $a = 2$, $b = 1$. The equation is then equivalent to the equation $z^2 = (r - 1)(3 - r)$. Thus $|z|$ is the mean proportional between $r - 1$

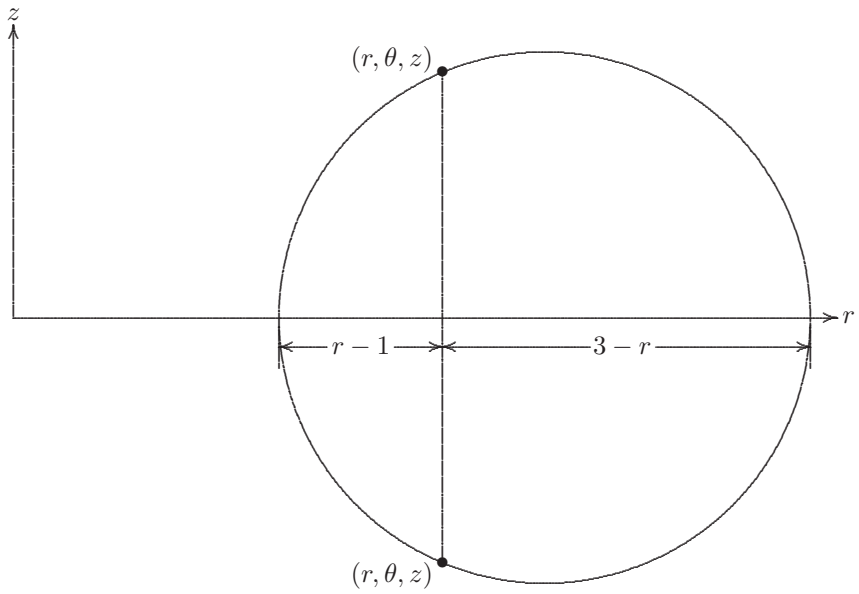


FIGURE 3. Cross-section of the torus, showing the two values of $z$ corresponding to a given value of $r$

and $3 - r$ and is defined—that is, the torus contains points $(r, \theta, z)$—if and only if $1 \leq r \leq 3$. The cross-section of the torus in the half-plane given by a fixed value of $\theta$ is shown in Fig. 3.

We define a chart $\psi_1$ as follows: The domain of $\psi_1$ is the set $U_1$ consisting of points $(r, \theta, z)$ for which $-\pi < \theta < \pi$ and $r > 1$, and its range $V_1$ is $\{\boldsymbol{x} = (x^1, x^2) : -\pi < x^1 < \pi\,,\ -\pi < x^2 < \pi\}$. Then

$$\begin{aligned} (x^1, x^2) = \psi_1(r, \theta, z) &= \left(\theta, 2\arctan\left(\frac{z}{r-1}\right)\right), \\ \psi_1^{-1}(x^1, x^2) &= \big(2 + \cos(x^2), x^1, \sin(x^2)\big)\,. \end{aligned}$$

In the first of these equations, it is the principal branch of the arctangent function that is meant, that is, the one whose values lie between $-\pi/2$ and $\pi/2$. This is an odd function, and since $r > 1$ on $U_1$, it follows that the sign of $x^2$ is the same as the sign of $z$, which in this range (from $-\pi$ to $\pi$) is the same as the sign of $\sin(x^2)$.

We now verify that the inverse function is the one stated in the second equality. Given that $x^2 = 2\arctan\big(z/(r-1)\big)$, we have $\tan(x^2/2) = z/(r-1) = \pm\sqrt{(3-r)/(r-1)}$, and hence $\sec^2(x^2/2) = 2/(r-1)$, which means that $\cos^2(x^2/2) = (r-1)/2$, and therefore $\sin^2(x^2/2) = (3-r)/2$, from which it follows that $\sin^2(x^2) = 4\sin(x^2/2)\cos(x^2/2) = (3-r)(r-1) = z^2$. But since $z$ and $\sin(x^2)$ have the same sign, we see that $z = \sin(x^2)$. It is immediately obvious that $\theta = x^1$. Finally, from what has already been shown, $\cos(x^2) = 2\cos^2(x^2/2) - 1 = (r-1) - 1 = r - 2$, and it follows that $r = 2 + \cos(x^2)$.

Similarly, we can define a chart $\psi_2$ by specifying that the domain $U_2$ consists of the points $(r,\theta,z)$ for which $0<\theta<2\pi$ and $r<3$, and the range $V_2$ is $\{\boldsymbol{y}:\, 0<y^1<2\pi\,,\ -\pi<y^2<\pi\}$. The mapping is then given by

$$\begin{aligned}(y^1,y^2)=\psi_2(r,\theta,z) &= \left(\theta, 2\arctan\left(\frac{z}{3-r}\right)\right),\\ \psi_2^{-1}(y^1,y^2) &= \big(2-\cos(y^2), y^1, \sin(y^2)\big)\,.\end{aligned}$$

The verification of the inverse relation here is left to the reader. We note that if $0<x^1<\pi$, then

$$(y^1,y^2)=\psi_2\circ\psi_1^{-1}(x^1,x^2)=\begin{cases}(x^1,\pi-x^2)\,, & \text{if } 0<x^2<\pi\,,\\ (x^1,-\pi-x^2)\,, & \text{if } -\pi<x^2<0\,,\end{cases}$$

while for $-\pi<x^1<0$,

$$(y^1,y^2)=\psi_2\circ\psi_1^{-1}(x^1,x^2)=\begin{cases}(x^1+2\pi,\pi-x^2)\,, & \text{if } 0<x^2<\pi\,,\\ (x^1+2\pi,-\pi-x^2)\,, & \text{if } -\pi<x^2<0\,,\end{cases}$$

The domain of $\psi_2\circ\psi_1^{-1}$ thus consists of the four disjoint open squares $(0,\pi)\times(0,\pi)$, $(0,\pi)\times(-\pi,0)$, $(-\pi,0)\times(0,\pi)$, $(-\pi,0)\times(-\pi,0)$, on each of which this function is analytic. Thus, $\mathbb{T}^2$ is an analytic manfiold. On each of the four components of this domain, the Jacobian determinant is defined as the formal expression

$$\det\left(\frac{\partial(y^1,y^2)}{\partial(x^1,x^2)}\right)=\det\begin{pmatrix}\frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2}\\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2}\end{pmatrix}=\det\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}=-1\,.$$

That this determinant is negative is a mere accident and could have been avoided simply by changing the sign of $x^2$ in the definition of $\psi_2$. If we had done so, the Jacobian would have been constantly 1.

**Example 4.4.** If, in addition to covering the manifold with the domains of charts that are $C^\infty$ compatible, we also require that there be at least one such covering for which the Jacobian of $\psi_1\circ\psi_2^{-1}$ defined for any two charts $\psi_1$ and $\psi_2$ is positive at every point , a manifold that meets the requirement is said to be *orientable*. Not every manifold is orientable, but we shall not be examining any non-orientable ones, with the single exception of the one we are now going to present, namely the *real projective plane* $\mathbb{P}^2$.

The real (as opposed to complex) projective plane $\mathbb{P}^2$ consists of pairs of antipodal points on the unit sphere $\mathbb{S}^2(1)$ in $\mathbb{R}^3$, that is, two-point sets of the form $\{\boldsymbol{\xi},-\boldsymbol{\xi}\}=\{(\xi^1,\xi^2,\xi^3),(-\xi^1,-\xi^2,-\xi^3)\}$. Since the points are on the unit sphere, we have $(\xi^1)^2+(\xi^2)^2+(\xi^3)^2=1$, and so in particular, at least one of the three coordinates in each pair is non-zero. Notice that a point in $\mathbb{P}^2$ is a *set* whose two elements are antipodal. It is *not* the *ordered pair* $(\boldsymbol{\xi},-\boldsymbol{\xi})$ which is different from $(-\boldsymbol{\xi},\boldsymbol{\xi})$. When we use this set notation for a point of $\mathbb{P}^2$, it does not matter which of the two antipodal points of $\mathbb{S}^2(1)$ that together make up the point of $\mathbb{P}^2$ is written first, since $\{a,b\}=\{b,a\}$ as sets. In general, when we define objects in terms of points of $\mathbb{P}^2$, actual computations require choosing one element of the pair that make up a given point, but the definitions are framed in such a way that it does not matter which of the two is chosen. A good example is provided by the set of three covering charts we are now going to define.

We first define a topology on $\mathbb{P}^2$ by saying that an open set $\widetilde{U}$ in $\mathbb{P}^2$ is a set of the form $\widetilde{U}=\big\{\{\boldsymbol{\xi},-\boldsymbol{\xi}\}\,:\,\boldsymbol{\xi}\in U\big\}$, where $U$ is an open set in $\mathbb{S}^2(1)$. Observing that

$\widetilde{U} = \widetilde{(-U)}$, where $-U = \{-\boldsymbol{\xi} : \boldsymbol{\xi} \in U\}$, one can see that this really is a topology. That is, one has only to remark that $\{\boldsymbol{\xi}, -\boldsymbol{\xi}\} \in \widetilde{U} \Leftrightarrow \boldsymbol{\xi} \in \big(U \cup (-U)\big)$.

We can cover $\mathbb{P}^2$ by three open sets $\widetilde{U}_1 = \big\{\{\boldsymbol{\xi}, -\boldsymbol{\xi}\} : \xi^1 \neq 0\big\}$, $\widetilde{U}_2 = \big\{\{\boldsymbol{\xi}, -\boldsymbol{\xi}\} : \xi^2 \neq 0\big\}$, and $\widetilde{U}_3 = \big\{\{\boldsymbol{\xi}, -\boldsymbol{\xi}\} : \xi^3 \neq 0\big\}$, and then define charts $\psi_i : U_i \to D$, where $D = \{(y^1, y^2) : (y^1)^2 + (y^2)^2 < 1$ as follows:

$$\begin{aligned}
\psi_1\big(\{\boldsymbol{\xi}, -\boldsymbol{\xi}\}\big) &= \operatorname{sgn}(\xi^1)(\xi^2, \xi^3)\,, \\
\psi_2\big(\{\boldsymbol{\xi}, -\boldsymbol{\xi}\}\big) &= \operatorname{sgn}(\xi^2)(\xi^1, \xi^3)\,, \\
\psi_3\big(\{\boldsymbol{\xi}, -\boldsymbol{\xi}\}\big) &= \operatorname{sgn}(\xi^3)(\xi^1, \xi^2)\,.
\end{aligned}$$

Here $\operatorname{sgn}(\xi) = \xi/|\xi|$ and is defined as long as $\xi \neq 0$. It makes sense even for complex numbers $\xi$, where $\operatorname{sgn}(\xi)$ becomes a point on the unit circle in the complex plane. However, we are dealing here with the real projective space $\mathbb{P}^2$, and accordingly, $\operatorname{sgn}(\xi)$ is either $+1$ (when $\xi > 0$) or $-1$ (when $\xi < 0$). Notice that it does not matter whether we use $\boldsymbol{\xi}$ or $-\boldsymbol{\xi}$ in this definition, since, for example, $\operatorname{sgn}(-\xi^1)(-\xi^2, -\xi^3) = \operatorname{sgn}(\xi^1)(\xi^2, \xi^3)$. The inverses of these mappings are

$$\begin{aligned}
\psi_1^{-1}(y^1, y^2) &= \Big\{\big(\sqrt{1 - (y^1)^2 - (y^2)^2}, y^1, y^2\big), \big(-\sqrt{1 - (y^1)^2 - (y^2)^2}, -y^1, -y^2\big)\Big\}\,, \\
\psi_2^{-1}(y^1, y^2) &= \Big\{\big(y^1, \sqrt{1 - (y^1)^2 - (y^2)^2}, y^2\big), \big(-y^1, -\sqrt{1 - (y^1)^2 - (y^2)^2}, -y^2\big)\Big\}\,, \\
\psi_3^{-1}(y^1, y^2) &= \Big\{\big(y^1, y^2, \sqrt{1 - (y^1)^2 - (y^2)^2}\big), \big(-y^1, -y^2, -\sqrt{1 - (y^1)^2 - (y^2)^2}\big)\Big\}\,.
\end{aligned}$$

As a result, we have the following compositions:

$$\begin{aligned}
\psi_2 \circ \psi_1^{-1}(y^1, y^2) &= \operatorname{sgn}(y^1)\big(\sqrt{1 - (y^1)^2 - (y^2)^2}, y^2\big)\,, \\
\psi_1 \circ \psi_2^{-1}(y^1, y^2) &= \operatorname{sgn}(y^1)\big(\sqrt{1 - (y^1)^2 - (y^2)^2}, y^2\big)\,, \\
\psi_3 \circ \psi_1^{-1}(y^1, y^2) &= \operatorname{sgn}(y^2)\big(\sqrt{1 - (y^1)^2 - (y^2)^2}, y^1\big)\,, \\
\psi_1 \circ \psi_3^{-1}(y^1, y^2) &= \operatorname{sgn}(y^1)\big(y^2, \sqrt{1 - (y^1)^2 - (y^2)^2}\big)\,, \\
\psi_2 \circ \psi_3^{-1}(y^1, y^2) &= \operatorname{sgn}(y^2)\big(y^1, \sqrt{1 - (y^1)^2 - (y^2)^2}\big)\,, \\
\psi_3 \circ \psi_2^{-1}(y^1, y^2) &= \operatorname{sgn}(y^2)\big(y^1, \sqrt{1 - (y^1)^2 - (y^2)^2}\big)\,.
\end{aligned}$$

The Jacobians of these six mappings are $\big(1 - (y^1)^2 - (y^2)^2\big)^{-1/2}$ times respectively $-y^1$, $-y^1$, $y^2$, $y^1$, $-y^2$, and $-y^2$. Each of them has a domain that consists of the open unit disk $(y^1)^2 + (y^2)^2 < 1$ with one of the axes removed. (It is the $y^2$-axis for the first, second, and fourth mappings and the $y^1$-axis for the third, fifth and sixth mappings. Thus, none of the mappings is defined at points where these Jacobians would be 0.) All six of them are defined in the open unit disk with the two axes removed, and it is apparent that none of the Jacobians is of constant sign. Reversing the sign of one of the coordinates in any or all of the mappings would not procure a system in which the Jacobians were all positive. Notice that $\mathbb{P}^2$ is *not* covered by the domains of any two of these three charts. We need all three.

**1.3. Functions on a manifold.** Let $M$ be a manifold and $P$ a point of $M$. A real-valued function $f$ whose domain is $M$ is defined to be continuous at $P$ if for each chart $\psi : U \to V$ at $P$, the composite function $f \circ \psi^{-1} : V \to \mathbb{R}^1$ is a continuous function. It suffices for this condition to hold for any one chart $\psi_1 : U_1 \to V_1$, since if $\psi_2 : U_2 \to V_2$ is a second chart at $P$, then on the (non-empty) open set $\psi_2(U_1)$

the function $f \circ \psi_2^{-1} = f \circ \psi_1^{-1} \circ (\psi_1 \circ \psi_2^{-1})$ is also continuous. This function is *not* a "non-existent" function, since $U_1 \cap U_2$ is non-empty (it contains $P$).

If $M$ is a $C^r$ manifold $0 \leq r \leq \omega$, it is reasonable to say that a real-valued function is a $C^k$ function on a neighborhood of $P$ for $k \leq r$ if $f \circ \psi^{-1}$ is a $C^k$ function for some (or every) chart $\psi$ at $P$. (By convention, a $C^\omega$ function is an analytic function.)

**1.4. Partitions of unity.** We are accustomed to carrying out the operation of integration in Euclidean space $\mathbb{R}^n$. Since a manifold looks locally like $\mathbb{R}^n$, we would like to be able to integrate over a manifold. This is easy to do if we want to integrate only over a small piece of the manifold, all of which is contained in the domain of a single chart. The chart enables us to transfer the integration to part of an open set in $\mathbb{R}^n$. But what do we do if we need to integrate over the entire manifold? The function (actually differential form) we wish to integrate needs to be written as a sum of functions, each of which is zero outside the domain of some chart. We can then integrate each of the latter functions and add their integrals. The technique for doing this involves the following considerations.

Let $\{U_\alpha\}$ be an open covering of a topological space $X$. A collection $\mathfrak{F}$ of continuous functions $f_\beta : X \to [0,1]$ is called a *partition of unity subordinate to the covering* $\{U_\alpha\}$ if it has the following properties:

1. For each function $f_\beta \in \mathfrak{F}$, there is an open set $U_\alpha$ in the covering such that $f(x) = 0$ for all $x \notin U_\alpha$. (This is the reason for the phrase *subordinate to the covering* $\{U_\alpha\}$.)
2. At each point $x \in X$, there is a neighborhod $N_x$ of $x$ and a finite set of indices $B_x = \{\beta_1, \ldots, \beta_n\}$ such that $f_\beta(y) = 0$ for all $y \in N_x$ unless $\beta \in B_x$.
3. For each $x \in X$ and all $y \in N_x$,
$$\sum_{j=1}^{n} f_{\beta_j}(y) = 1\,.$$
This is the reason for calling the collection $\mathfrak{F}$ a *partition of unity*.

Given a set of functions $\mathfrak{F}'$ having the first two of these properties and the additional property that
$$\sum_{\beta \in B_x} f_\beta(x) > 0\,,$$
it makes sense to define the sum of all the functions in $\mathfrak{F}'$ as the function
$$F(x) = \sum_{\beta \in B_x} f_\beta(x).$$

Since $F(y) = f_{\beta_1}(y) + \cdots + f_{\beta_n}(y)$ for $y \in N_x$, we see that $F$ coincides with a continuous function on a neighborhood of each point $x$, and hence is a continuous function that assumes only positive values. If we then define $\mathfrak{F}$ to be the set of all functions $f_\beta/F$ for $f \in \mathfrak{F}'$, we see easily that $\mathfrak{F}$ is a partition of unity.

Thus to construct a partition of unity, we need to get a collection having the properties of $\mathfrak{F}'$. We want our partitions to consist of functions that are actually $C^\infty$ functions. The key to getting such a partition of unity is the following lemma.

**Lemma 4.1.** *The function $f_r : \mathbb{R}^1 \to \mathbb{R}^1$ given by*
$$f_r(x) = \begin{cases} e^{\frac{1}{x^2 - r^2}}\,, & \text{if } |x| < r \\ 0\,, & \text{if } |x| \geq r\,, \end{cases}$$

*is a $C^\infty$ function.*

PROOF. Since

$$\frac{1}{x^2-r^2} = \Big(\frac{1}{x-r} - \frac{1}{x+r}\Big)/(2r)\,,$$

we see that for $0 \le x < r$ we have

$$f_r(x) = \varphi_r(x) e^{\frac{1}{2r(x-r)}}\,.$$

where $\varphi_r(x)$ is a $C^\infty$ function on $(-r, +\infty)$. It is obvious that $f_r(x) \to 0$ as $x \uparrow r$, and by induction we can show that the same is true for all derivatives $f_r^{(n)}(x)$. This is merely a matter of L'Hospital's rule, since one can show easily that

$$\lim_{x\uparrow r} f_r^{(n)}(x) = \lim_{x\uparrow r} \frac{p_n\big(\frac{1}{r-x}\big)}{e^{1/(2r(r-x))}} = \lim_{t\to+\infty} \frac{p_n(t)}{e^{t/(2r)}}\,,$$

where $t = 1/(r-x)$ and $p_n(t)$ is a polynomial of degree $2n$. It then follows by induction on $n$ and the mean-value theorem that $f_r^{(n-1)}(x)$ has a left-hand derivative at $x = r$ equal to 0. Since the right-hand derivative at that point is also 0 (because $f_r(x) \equiv 0$ for $x \ge r$) it follows that $f_r^{(n)}(r)$ exists and equals 0, completing the induction. Thus, $f_r(x)$ is a $C^\infty$ function on $(-r, +\infty)$.

The same argument applies at $x = -r$, so that $f_r(x)$ is a $C^\infty$ function on $(-\infty, \infty)$ whose values are in the range $(0, 1)$ if $|x| < r$ and 0 if $|x| \ge r$. (Obviously, the maximum value of $f_r(x)$ is $f_r(0) = e^{-1/r^2}$, which is less than 1.) □

We get a $C^\infty$ function $f_r : \mathbb{R}^n \to [0, 1]$ by setting

$$f_r(\boldsymbol{x}) = \begin{cases} e^{\frac{1}{|\boldsymbol{x}|^2-r^2}}\,, & \text{if } |\boldsymbol{x}| < r\,, \\ 0\,, & \text{if } |\boldsymbol{x}| \ge r\,. \end{cases}$$

We now have a $C^\infty$ function $f_{r,\boldsymbol{x}_0}$ that is positive inside the open ball $B_r(\boldsymbol{x}_0)$ and 0 elsewhere, namely $f_{r,\boldsymbol{x}_0}(\boldsymbol{x}) = f_r(\boldsymbol{x} - \boldsymbol{x}_0)$.

With the very useful family of functions $f_{r,\boldsymbol{x}_0}(\boldsymbol{x})$ we can now construct other useful $C^\infty$ functions on $\mathbb{R}^n$. For example, if $C$ is a compact set and $E$ a closed set disjoint from $C$, for each point $\boldsymbol{x} \in C$, there is a ball $B_r(\boldsymbol{x})$ that is disjoint from $E$. Some finite set of these, say $B_{r_1}(\boldsymbol{x}_1), \ldots, B_{r_n}(\boldsymbol{x}_n)$ covers the compact set $C$. Then the function $f(\boldsymbol{x}) = f_{r_1,\boldsymbol{x}_1}(\boldsymbol{x}) + \cdots + f_{r_n,\boldsymbol{x}_n}(\boldsymbol{x})$ is positive on an open set containing $C$ and 0 on $E$.

Refining this argument still further, we get the following important result:

**Theorem 4.1.** *Let $U$ be any open subset of $\mathbb{R}^n$. There exists a $C^\infty$ function $f : \mathbb{R}^n \to \mathbb{R}^1$ that is positive on $U$ and 0 on $\mathbb{R}^n \setminus U$.*

PROOF. If $U = \varnothing$, $f(\boldsymbol{x}) \equiv 0$ will do, and no other function will. If $U = \mathbb{R}^n$, $f(\boldsymbol{x}) \equiv 1$ will do, as will many others. Now let $U$ be a non-empty open set for which $\mathbb{R}^n \setminus U$ is also non-empty.

For each positive integer $n$, Let $C_n$ be the compact subset defined as follows:

$$C_n = \{\boldsymbol{x} \in \mathbb{R}^n : |\boldsymbol{x}| \le n\,,\, B_{1/n}(\boldsymbol{x}) \subseteq U\}\,.$$

We claim $C_n$ is a compact subset of $U$. That it is a subset of $U$ is obvious from its definition, and the first condition in that definition assures that it is bounded. By the Heine–Borel theorem proved in Appendix 3, it now suffices to show that $C_n$ is closed. To that end, suppose $\boldsymbol{y} \in \mathbb{R}^n \setminus C_n$. Then, either $|\boldsymbol{y}| > n$ or $B_{1/n}(\boldsymbol{y})$

contains a point $\boldsymbol{z}$ belonging to $\mathbb{R}^n \setminus U$. In the first case $B_r(\boldsymbol{y}) \subseteq \mathbb{R}^n \setminus C_n$ for $r < |\boldsymbol{y}| - n$. In the second case, we claim that $B_{1/n}(\boldsymbol{x})$ contains a point of $\mathbb{R}^n \setminus U$ (namely $\boldsymbol{z}$) if $|\boldsymbol{y} - \boldsymbol{x}| < 1/n - |\boldsymbol{z} - \boldsymbol{y}| = \delta$. Indeed, $|\boldsymbol{z} - \boldsymbol{x}| \leq |\boldsymbol{z} - \boldsymbol{y}| + |\boldsymbol{y} - \boldsymbol{x}| < |\boldsymbol{z} - \boldsymbol{y}| + 1/n - |\boldsymbol{z} - \boldsymbol{y}| = 1/n$, so that $\boldsymbol{z} \in B_{1/n}(\boldsymbol{x}) \setminus U$. Thus $B_\delta(\boldsymbol{y}) \subseteq \mathbb{R}^n \setminus C_n$ in this case, and so in either case, $\mathbb{R}^n \setminus C_n$ is open. Therefore $C_n$ is closed and bounded and hence compact.

The sets $C_n$ form an increasing sequence of subsets of $U$ whose union is $U$. We now claim that the open set $U_n$, defined as the union of the open balls $B_{1/(n(n+1))}(\boldsymbol{y})$ as $\boldsymbol{y}$ ranges over $C_n$, is contained in $C_{n+1}$, and hence in the interior of $C_{n+1}$. For, if $\boldsymbol{y} \in C_n$ and $|\boldsymbol{x} - \boldsymbol{y}| \leq 1/(n(n+1))$, then $|\boldsymbol{x}| \leq |\boldsymbol{y}| + |\boldsymbol{x} - \boldsymbol{y}| < n + 1/(n(n+1)) < n+1$. Also, $B_{1/(n+1)}(\boldsymbol{x}) \subseteq B_{1/n}(\boldsymbol{y}) \subseteq U$ since if $\boldsymbol{z} \in B_{1/(n+1)}(\boldsymbol{x})$, then $|\boldsymbol{z} - \boldsymbol{y}| \leq |\boldsymbol{z} - \boldsymbol{x}| + |\boldsymbol{x} - \boldsymbol{y}| < 1/(n+1) + 1/(n(n+1)) = 1/n$, that is, $B_{1/(n+1)}(\boldsymbol{x}) \subseteq B_{1/n}(\boldsymbol{y}) \subseteq U$ and so $\boldsymbol{x} \in C_{n+1}$. These two conditions assure that $B_{1/(n(n+1))}(\boldsymbol{y}) \subseteq C_{n+1}$, and hence $\boldsymbol{y} \in \operatorname{int}(C_{n+1})$.

Now consider the sequence of compact sets $D_1 = C_1$, $D_2 = C_2 \setminus \operatorname{int}(C_1), \ldots$, $D_n = C_n \setminus \operatorname{int}(C_{n-1}), \ldots$. The union of these sets is the same as the union of the sets $C_n$. If $k > n+1$, then $D_k$ is disjoint from the open set $V_n$ that is the union of the balls $B_{1/(n(n+1))}(\boldsymbol{y})$ as $\boldsymbol{y}$ ranges over $D_n$. Thus, $D_n \subseteq V_n$ and $V_n \cap D_k = \varnothing$ if $k > n+1$. For $V_n$, as just shown, is contained in the interior of $C_{n+1}$, and $D_k$ is disjoint from the interior of $C_{k-1}$, and so also disjoint from the interior of $C_{n+1}$.

Now let $W_n$ be the intersection of $U$ with the union of the balls $B_{1/(2n(n+1))}(\boldsymbol{y})$ as $\boldsymbol{y}$ ranges over $D_n$. Then no point belongs to more than two of the sets $W_n$. A point can belong to both $W_n$ and $W_{n+1}$, but not to $W_k$ if $k \geq n+2$.

As above, we can form a $C^\infty$ function $f_n(\boldsymbol{x})$ that is positive on $D_n$ and 0 outside of $W_n$. Then the infinite sum

$$f(\boldsymbol{x}) = \sum_{n=1}^{\infty} f_n(\boldsymbol{x})$$

is a finite sum in a neighborhood of each point of $U$, since such a point belongs to $D_n$ for some $n$, and hence not to $D_k$ if $k \geq n+2$. Since every point of $U$ belongs to some $D_n$, the sum is positive on $U$ and 0 outside of $U$, since all the $f_n$ vanish outside of $U$. □

**Corollary 4.1.** *Let $M$ be a paracompact $C^\infty$ manifold covered by a collection of open sets $\mathfrak{U}$. There exists a $C^\infty$ partition of unity on $M$ subordinate to $\mathfrak{U}$.*

PROOF. Let $\{\psi_\alpha\}\alpha \in A$ be a collection of charts whose domains $\{U_\alpha\}_{\alpha \in A}$ also cover $M$. By passing, if necessary, to the collection $\mathfrak{U}'$ consisting of all intersections $U \cap U_\alpha$, where $U \in \mathfrak{U}$ and $\alpha \in A$, we can assume that each set $U \in \mathfrak{U}$ is contained in some $U_\alpha$. Now consider the collection $\mathfrak{V}$ consisting of of open sets $V$ having the property that $V^c \subseteq U$ for some $U \in \mathfrak{U}$. Since each point $P \in M$ has a compact neighborhood contained in some set $U$, the collection $\mathfrak{V}$ also covers $M$. Since $M$ is paracompact, there exists a locally finite refinement $\mathfrak{V}'$ of $\mathfrak{V}$. For each set $V \in \mathfrak{V}'$, we construct a $C^\infty$ mapping $h_V : M \to \mathbb{R}$ as follows.

Choose and keep fixed an index $\alpha \in A$ such that $V^c \subseteq U_\alpha$. This is possible, since there exists $U \in \mathfrak{U}$ such that $V^c \subseteq U$, and by our reduction there exists $\alpha \in A$ such that $U \subseteq U_\alpha$.

Next, let $g_V : \mathbb{R}^n \to \mathbb{R}$ be a $C^\infty$ function that is positive on the set $\psi_\alpha(V)$ and zero outside this set. Since $\psi_\alpha$ is a homeomorphism, $\psi_\alpha(V)^c = \psi_\alpha(V^c)$ which

is a closed subset of $\psi_\alpha(U_\alpha)$. In particular, if $\boldsymbol{y} \notin \psi_\alpha(U_\alpha)$, then $\mathbb{R}^n \setminus \psi_\alpha(V)^c$ is a neighborhood of $\boldsymbol{y}$ on which $g_V$ is identically zero.

We now let

$$h_V(Q) = \begin{cases} g_V\big(\psi_\alpha(Q)\big)\,, & \text{if } \ Q \in U_\alpha\,, \\ 0\,, & \text{if } \ Q \notin U_\alpha\,. \end{cases}$$

The function $h_V(Q)$ is obviously $C^\infty$ on $U_\alpha$, since by definition, this means only that $h_V \circ \psi_\alpha^{-1} = g_V$ is a $C^\infty$ function on $\psi_\alpha(U_\alpha)$. But if $Q \notin U_\alpha$, then $M \setminus V^c$ is an open neighborhood of $Q$ on which $h_V(Q) \equiv 0$, and so $h_V$ is a $C^\infty$ function in a neighborhood of every point. We observe that $h_V(Q)$ is zero for $Q \notin V$ and positive for $Q \in V$.

We now define a positive-valued $C^\infty$ function $H : M \to \infty$ by setting

$$H(Q) = \sum_{V \in \mathfrak{V}'} h_V(Q)\,.$$

Since each point $Q$ has a neighborhood that intersects only a finite number of the sets $V \in \mathfrak{V}'$, this sum is a finite sum in a neighborhood of each point, and since each term is a $C^\infty$ function, $H(Q)$ is also a $C^\infty$ function. Since each $Q \in M$ belongs to at least one of the sets $V$, the sum is positive at every point.

If we now set $f_V = h_V/H$, then the collection

$$\{f_V\}_{V \in \mathfrak{V}'}$$

is a partition of unity on $M$ subordinate to the given open covering of $M$. □

**Remark 4.1.** It is easy to see that the refinement $\mathfrak{V}'$ of $\mathfrak{V}$ is actually a subcovering. For if $W \subseteq V$, where $V \in \mathfrak{V}$, then $V^c \subseteq U_\alpha$, and hence $W^c \subseteq V^c \subseteq U_\alpha$ also, that is, $W \in \mathfrak{V}$.

**Remark 4.2.** If the covering consists of a finite or countable number of sets $U_1, U_2, \ldots,$, then we can group the functions in the partition of unity, denoting by $f_1$, the sum of all those whose support (a *closed* set, remember) is contained in $U_1$, then by $f_2$ all those not summed as part of $f_1$ and whose support is contained in $U_2$, and so on. In this way, we can associate a single function $f_i$ with each set in the covering. If the covering is minimal, then none of these functions will be identically zero.

**Remark 4.3.** The preceding proof is a pure existence proof, in which we invoke all the godlike powers we have endowed ourselves with to define objects in terms of other objects for which we have no explicit construction. Such a proof works for abstract manifolds, that is, for the *class* of manifolds, or for a manifold about which nothing specific is given. On any particular manifold, whose definition is explicit, a partition of unity can often be constructed explicitly, as the following example shows.

**Example 4.5.** Let $M = \mathbb{S}^n(r)$ be the $n$-dimensional sphere of radius $r$ in $\mathbb{R}^{n+1}$. The simplest possible partition of unity subordinate to the domains of the two stereographic projections $\psi_N$ and $\psi_S$, is constructed as follows:

$$f_N(\boldsymbol{\xi}) = \begin{cases} e^{-|\frac{\psi_N(\boldsymbol{\xi})}{r}|^2}\,, & \text{if } \ \boldsymbol{\xi} \neq N = (0, 0, \ldots, 0, r)\,, \\ 0\,, & \text{if } \ \boldsymbol{\xi} = N\,. \end{cases}$$

It is obvious that $f_N \circ \psi_N^{-1}(\boldsymbol{x}) = e^{-|\frac{\boldsymbol{x}}{r}|^2}$ is a $C^\infty$ function for all $\boldsymbol{x} \in \mathbb{R}^n$, so that $f_N$ is $C^\infty$ provided it is also $C^\infty$ at $N$. At that point, we need to verify that $f_N \circ \psi_S^{-1}(\boldsymbol{x})$ is $C^\infty$. But

$$f_N \circ \psi_S^{-1}(\boldsymbol{x}) = \begin{cases} e^{-\frac{r^2}{|\boldsymbol{x}|^2}}, & \text{if } \boldsymbol{x} \neq \mathbf{0}, \\ 0, & \text{if } \boldsymbol{x} = \mathbf{0}. \end{cases}$$

and, as we know, this function and all of its partial derivatives vanish at $\boldsymbol{x} = \mathbf{0} = \psi_S(N)$.

Since $0 < f_N(\boldsymbol{\xi}) < 1$ when $\boldsymbol{\xi}$ is neither $N$ nor $S$, and $f_N(N) = 0$, $f_N(S) = 1$, it follows easily that $f_S(\boldsymbol{\xi}) = 1 - f_N(\boldsymbol{\xi})$ is a $C^\infty$ function that is zero at $S$ and positive elsewhere. Obviously, then $f_N(\boldsymbol{\xi}) + f_S(\boldsymbol{\xi}) \equiv 1$, and we have a partition of unity subordinate to the given covering of $\mathbb{S}^n(r)$.

**1.5. The topology generated by charts.** It was mentioned in the introduction to this chapter that manifolds are always locally compact. In fact, the definition of a manifold does not seem to involve the topology on the manifold in any essential way. What is important is that the charts be one-to-one mappings such that the composition $\psi_2 \circ \psi_1^{-1}$ has a certain amount of smoothness. But this mapping has domain and range in $\mathbb{R}^n$, and would have the same smoothness properties regardless of the smoothness of $\psi_1$ and $\psi_2$ in terms of the topology of $M$. We could, if we wished, start with these mappings and then define a topology on $M$ just strong enough to make them all continuous. How far can we get if we start with a non-empty set $M$ and a collection $\{U_\alpha\}_{\alpha \in A}$, of non-empty subsets $U_\alpha$ of $M$ that cover $M$ as $\alpha$ ranges over the index set $A$? We cannot yet call the sets $U_\alpha$ open sets since we are not assuming any topology on $M$. With each $\alpha \in A$, we also have a one-to-one mapping $\psi_\alpha : U_\alpha \to V_\alpha$, where $V_\alpha$ is an open subset of $\mathbb{R}^n$ *onto* which $\psi_\alpha$ maps $U_\alpha$. The crucial property that imposes a manifold structure on $M$ is that for each pair of indices $\alpha \in A$, $\beta \in A$, the mapping $\psi_\alpha \circ \psi_\beta^{-1}$ is continuous. We now impose the weakest topology that makes all the charts continuous. That is, open sets are unions of sets $\psi_\alpha^{-1}(V_\alpha)$, where each $V_\alpha$ is an open subset of $\mathbb{R}^n$. Certainly, the domain $U_\alpha$ of $\psi_\alpha$ is now an open set, since the range of this mapping is open in $\mathbb{R}^n$. Since one-to-one mappings preserve all set operations, all the mappings $\psi_\alpha$ and their inverses $\psi_\alpha^{-1}$ are now continuous, that is, they are all homeomorphisms, and we can now get a complete manifold structure by adjoining all possible homeomorphisms from open subsets of $M$ into $\mathbb{R}^n$.

What properties will this topology have? As already stated, it will certainly be a locally compact topology, for example, since open sets in $\mathbb{R}^n$ are locally compact. Will it be $\sigma$-compact? The answer is yes if we have only a countable set of charts to begin with, since the domain of each chart is $\sigma$-compact. Will $M$ be paracompact? As we saw in Appendix 2, it will be if it is a Hausdorff space. Thus we come to the crucial "Goldilocks" question: Will this be a Hausdorff topology? Unfortunately, the answer to that question is, in general, no, as we shall now prove. First, let us state a few more details about the topology we have defined, to make it completely clear that it is a locally compact topology on $M$

We have defined the topology on $M$ by taking as a sub-basis the collection of sets $\varphi_\alpha^{-1}(V)$, where $V$ is an open subset of $\mathbb{R}^n$ and $\alpha \in A$. The open sets in this topology are all unions of finite intersections of sets of this type. In other words, a set $U \subseteq M$ is open if and only if for some index set $C \subseteq A$ and each $\gamma \in C$, there

is a finite, non-empty collection of indices $\alpha_{\gamma 1}, \ldots, \alpha_{\gamma n_\gamma}$ from the original index set $A$ and open subsets $W_{\gamma j}$ of $\mathbb{R}^n$, $j = 1, \ldots, n_\gamma$, such that

$$U = \bigcup_{\gamma \in C} \bigcap_{j=1}^{n_\gamma} \psi_{\alpha_{\gamma j}}^{-1}(W_{\gamma_j})\,.$$

We note that, since $\psi_{\alpha_{\gamma_j}}^{-1}(W_{\gamma_j}) = \psi_{\alpha_{\gamma_j}}^{-1}(W_{\gamma_j} \cap V_{\alpha_{\gamma_j}})$, we may assume that $W_{\gamma_j} \subseteq V_{\alpha_{\gamma_j}}$ in this definition.

With this definition, obviously, each mapping $\psi_\alpha$ is continuous. We would like to show that it is also an open mapping, so that the image of any open set is open. Consider the set $U$ in the definition given above. Since $\psi_\alpha$ is one-to-one, it preserves set operations, and therefore

$$\begin{aligned}
\psi_\alpha(U \cap U_\alpha) &= \psi_\alpha\Big(U_\alpha \cap \Big(\bigcup_{\gamma \in C} \bigcap_{j=1}^{n_\gamma} \big(\psi_{\alpha_{\gamma_j}}^{-1}(W_{\gamma_j})\big)\Big)\Big) \\
&= \psi_\alpha\Big(\bigcup_{\gamma \in C} \bigcap_{j=1}^{n_\gamma} U_\alpha \cap \big(\psi_{\alpha_{\gamma_j}}^{-1}(W_{\gamma_j})\big)\Big) \\
&= \bigcup_{\gamma \in C} \bigcap_{j=1}^{n_\gamma} \psi_\alpha\big(U_\alpha \cap \psi_{\alpha_{\gamma_j}}^{-1}(W_{\gamma_j})\big) \\
&= \bigcup_{\gamma \in C} \bigcap_{j=1}^{n_\gamma} \psi_\alpha \circ \psi_{\alpha_{\gamma_j}}^{-1}(W_{\gamma_j})\,.
\end{aligned}$$

Now $W_{\gamma_j}$ is an open set in $\mathbb{R}^n$ and $\psi_\alpha \circ \psi_{\alpha_{\gamma_j}}^{-1}$, as noted above, is a homeomorphism. The image $\psi_\alpha \circ \psi_{\alpha_{\gamma_j}}^{-1}(W_{\gamma_j})$ is therefore an open set in $\mathbb{R}^n$. It follows that $\psi_\alpha(U_\alpha \cap U)$ is a union of finite intersections of sets that are open in $\mathbb{R}^n$, and hence is itself open. Thus, each $\psi_\alpha$ is a homeomorphism. (Of course, all these statements are likely to be trivial, as $U \cap U_\alpha$ is likely to be the empty set!)

It is immediate that each point $P \in M$ has a compact neighborhood, since $\psi(P)$ has a compact neighborhood $N$, and $\psi^{-1}(N)$ being the image of a compact set under the continuous mapping $\psi^{-1}$ is compact. The point $P$ is in its interior, since it belongs to the open set $\psi^{-1}\big(\mathrm{int}\,(N)\big)$, which is contained in $\psi^{-1}(N)$. Thus, $M$ is locally compact. Since the range of each chart is $\sigma$-compact, the domain is also. Hence if the set of charts is countable, then $M$ itself is a countable union of compact sets.

That now being clear, it remains to show by example that the topology generated by the charts alone can be a non-Hausdorff topology. Here is an example, adapted from *Geometry of Manifolds*, by Richard L. Bishop and Richard J. Crittenden, (AMS Chelsea, 2001), pp. 5–6.

Let $M$ be the open interval $\big(-\frac{1}{4}, \frac{5}{4}\big)$ on the real line, and define two mappings $\varphi_1 : U_1 \to V_1$ and $\varphi_2 : U_2 \to V_2$, where $U_1 = M \setminus \{0\}$, $\varphi_1(x) = x$, $U_2 = \big(-\frac{1}{4}, 0\big] \cup \big(1, \frac{5}{4}\big)$, $\varphi_2(x) = \sin(2\pi x)$. Obviously, $V_1 = U_1 = \big(-\frac{1}{4}, 0\big) \cup \big(0, \frac{5}{4}\big)$, and it is easy to see that $\varphi_2$ is a one-to-one mapping of $U_2$ onto $(-1, 1)$. The mapping $\varphi_1 \circ \varphi_2^{-1}$ is defined as follows:

$$\varphi_1 \circ \varphi_2^{-1}(x) = \begin{cases} \arcsin(x)/(2\pi) & \text{if } -1 < x < 0\,, \\ 1 + \arcsin(x)/(2\pi) & \text{if } 0 < x < 1. \end{cases}$$

The mapping $\varphi_1 \circ \varphi_2^{-1}(x)$ is not defined at $x = 0$, and that is how it manages to be a $C^\infty$ mapping. Likewise $\varphi_2 \circ \varphi_1^{-1}(x) = \sin(2\pi x)$ if $x \neq 0$, and this mapping is not defined for $x = 0$.

If we impose only the weakest topology that makes these mappings homeomorphisms, then a neighborhood of any point of $M$ except 0 is simply an ordinary neighborhood of that point, while a basic neighborhood of 0 can be $\varphi_2^{-1}(-\varepsilon, \varepsilon) = \left\{\frac{1}{2\pi}\arcsin(x) \,:\, -\varepsilon < x \leq 0\right\} \cup \left\{1 + \frac{1}{2\pi}\arcsin(x)) \,:\, 0 < x < \varepsilon\right\}$. Thus every neighborhood of 1 intersects every neighborhood of 0. We shall try to avoid such pathologies and take it for granted that any structure we define generates a topology that is at least a Hausdorff topology. This is never difficult to achieve. For any manifold embedded in a Euclidean space $\mathbb{R}^k$ of any dimension, for example, the Hausdorff condition holds automatically.

**1.6. The full differentiable structure.** We now consider a base for a manifold structure on $M$ in which all the charts are $C^\infty$ functions. If we adjoin all possible homeomorphisms that are $C^\infty$ compatible with those in the base, we get a full manifold structure in which any two charts are $C^\infty$ compatible. That means that if $\psi$ is such a chart and has domain $T$, then $\psi(T \cap U_\alpha)$ and $\varphi_\alpha(T \cap U_\alpha)$ are both open sets in $\mathbb{R}^n$, and $\psi \circ \varphi_\alpha^{-1}$ maps the latter in a one-to-one $C^\infty$ way onto the former. In any actual computation, we have no need of anything but a basis for the differentiable structure. However, in theoretical arguments, it would be too much trouble to have to write continually, "if the chart $\psi$ is adjoined to the differentiable structure…". It is simpler just to assume that any admissible chart is already in the structure.

## 2. Differential Operators and The Tangent Space at a Point

Manifolds are introduced so that we can study vector fields on them. To do that, we need to be able to take derivatives. But since a manifold is not a Euclidean space, the derivatives must somehow be "transferred in" from the ranges of various charts and pieced together so as to get an operator that resembles a derivative. For this purpose, as we are about to see, the most convenient object by far is a manifold that is $C^\infty$ or $C^\omega$. Let us now explore the reasons why we do not consider $C^r$ manifolds for finite $r$ (although some people do!). This discussion is essentially an exposition of the contents of the paper "Tangent planes to a differentiable manifold," by W. F. Newns and A. G. Walker, *Journal of the London Mathematical Society*, 1956, 400–407.

**2.1. $C^0$ manifolds.** When we assume only continuity, we find that no notion of differentiation can be introduced that will apply to all continuous functions.

**Theorem 4.2.** *Let $X$ be a topological space and $x \in X$. Let $\mathfrak{A}$ be the algebra of all continuous real-valued functions on $X$, and let $L : \mathfrak{A} \to \mathbb{R}$ be a linear functional such that $L(f^2) = 2f(x)L(f)$ for all $f \in \mathfrak{A}$. Then $L(f) \equiv 0$.*

**Remark 4.4.** By using the polarization formula $uv = \big((u+v)/2\big)^2 - \big((u-v)/2\big)^2$, we find that $L(fg) = f(x)L(g) + g(x)L(f)$ for all $f$ and $g$ in $\mathfrak{A}$. A linear functional on an algebra of functions that has this property will be called a *derivation*.

Proof. We first note that $L(f) = 0$ if $f$ is a constant function. This is easily proved: If $f_1(y) \equiv 1$ for all $y$, then $L(f_1) = L(f_1^2) = 2f_1(x)L(f_1) = 2L(f_1)$,

and hence $L(f_1) = 0$. Then, for any constant $a$, if $f_a(y) \equiv a$ for all $y$, we have $L(f_a) = L(af_1) = aL(f_1) = 0$.

Next, we see that if $f(y) \geq 0$ for all $y$, and $f(x) = 0$, then $L(f) = 0$. We prove this by considering the function $g(y) = 1 + \sqrt{f(y)}$. On the one hand,

$$L(g^2) = 2g(x)L(g) = 2L\big(1 + \sqrt{f}\big) = 2L\big(\sqrt{f}\big)\,.$$

On the other hand,

$$L(g^2) = L\big(1 + 2\sqrt{f} + f\big) = 2L\big(\sqrt{f}\big) + L(f)\,.$$

By subtraction, $L(f) = 0$.

Next we see that if $f(x) = 0$, then $L(f) = 0$. For

$$L\big(|f| + f\big) = 0$$

and

$$L\big(|f| - f\big) = 0\,.$$

Hence,

$$L(f) = L\big(|f|\big) = -L(f)\,,$$

and so, $L(f) = 0$.

Finally, we write

$$L(f) = L\big(f - f(x)\big) + L\big(f(x)\big) = 0 + 0 = 0\,.$$

□

This proof carries over with no changes at all when the algebra of continuous functions on $X$ is replaced by the algebra of functions $f$ defined on a neighborhood of a point $P$ on a $C^0$ manifold $M$ and continuous at $P$. (The neighborhood on which $f$ is defined may depend on $f$.)

If we consider the smaller algebra $\mathfrak{A}$ of functions $f$ that are defined on a neighborhood of a point $P$ (again, the neighborhood may depend on $f$) and are such that $\frac{\partial \tilde{f}}{\partial y^i}$ exists at the point $y = \psi(P)$, where $\psi$ is a chart at $P$ and $\tilde{f} = f \circ \psi^{-1}$, then there are non-zero derivations $L : \mathfrak{A} \to \mathbb{R}$, for example $L(f) = 7\frac{\partial \tilde{f}}{\partial y^i}\big(\psi(P)\big)$. For convenience, we shall use the notation $\frac{\partial}{\partial x^i}$ for this derivation. Thus,

$$\frac{\partial f}{\partial x^i} = \frac{\partial \tilde{f}}{\partial y^i}\big(\psi(P)\big)\,.$$

**2.2.** $C^r$ **manifolds,** $0 < r < \infty$**.** To see what happens for finite positive values of $r$, we need first a simple preliminary result.

**Lemma 4.2.** *Let $r$ be a positive integer, and let $f$ and $g$ be $C^r$ functions on a neighborhood of $P$ for which $f(P) = 0 = g(P)$, then all partial derivatives of $fg$ of order $r + 1$ exist at $P$.*

Proof. This lemma is an easy consequence of the Leibniz's rule for the higher-order derivatives of a product, whereby if $h$ is a partial derivative of $fg$ of order $r$, then

$$h = fg_r + gf_r + \sum_{k=1}^{r-1} f_k g_{r-k}\,,$$

where $f_k g_{r-k}$ denotes a linear combination of products each of which is a partial derivative of $f$ of order $k$ multiplied by a partial derivative of $g$ of order $r-k$ (if $r=1$, this last sum is zero) and $f_r$ and $g_r$ are partial derivatives of order $r$.

It then follows from the fact that $f$ and $g$ are $C^r$ functions that all the first-order partial derivatives of the terms in the sum exist and are continuous at $P$. We now need only show that all the first-order partial derivatives of $fg_r + gf_r$ at $P$ exist. Again, this is easy to do. If $y = (y^1, \dots, y^n)$, where $y^j = 0$ for $j \neq i$, then

$$\frac{\partial(fg_r + gf_r)}{\partial x^i} = \lim_{y^i \to 0} \left( \frac{\tilde{f}(y)}{y^i} \tilde{g}_r(y) + \frac{\tilde{g}(y)}{y^i} \tilde{f}_r(y) \right) = \frac{\partial \tilde{f}}{\partial y^i} \tilde{g}_r(0) + \frac{\partial \tilde{g}}{\partial y^i} \tilde{f}_r(0)\,,$$

the partial derivative on the left being evaluated at $P$ and those on the right at $\psi(P) = 0$. □

We now prove the following theorem, whose statement and proof we have adapted and edited but whose basic structure and ideas can be found in the paper of Newns and Walker cited above:

**Theorem 4.3.** *Let $\psi$ be a chart at a point $P$ on a $C^r$ manifold, where $0 < r < \infty$. Let $\mathfrak{A}$ be the algebra of functions defined on a neighborhood of the point $P$—the domain of definition may depend on the function—that are continuous along with all partial derivatives of order up to and including $r$, and let $V$ be the vector space of derivations $L$ of $\mathfrak{A}$ at the point $P$. The dimension of $V$ is $\mathfrak{c}$, the cardinality of the continuum.*

Proof. Fix a chart $\psi$ at $P$, and assume without any loss of generality that $\psi(P) = 0$. Let $\mathfrak{A}_1$ be the subalgebra of $\mathfrak{A}$ consisting of functions $f$ for which all partial derivatives of $f$ of order $r+1$ exist[3] at $P$, that is, the corresponding partial derivatives of $\tilde{f} = f \circ \psi^{-1}$ of order $r+1$ exist at $\psi(P) = 0$.

For all $\eta \in (0,1)$, let

$$f_\eta(x) = |\psi(x)|^{r+\eta}\,,$$

that is, $\tilde{f}_\eta(y) = f_\eta\big(\psi^{-1}(y)\big) = |y|^{r+\eta}$.

It is easy to prove by induction that there are constants $C^s_{i_1,\dots,i_n}$ defined for non-negative integers $i_1, \dots, i_n, s$ satisfying $0 \leq i_1 + \cdots + i_n \leq s$, such that if $g(y)$ is any partial derivative of $\tilde{f}_\eta(y)$ of order $s$, then for $y \neq 0$

$$g(y) = \sum_{0 \leq i_1 + \cdots + i_n \leq s} C^s_{i_1,\dots,i_n} |y|^{r+\eta-2(i_1+\cdots+i_n)} (y^1)^{i_1} \cdots (y^k)^{i_n}\,,$$

and in fact, by an easy induction,

$$\frac{\partial g}{\partial y^k} = \sum_{0 \leq i_1 + \cdots + i_n \leq s+1} C^{s+1}_{i_1,\dots,i_n} |y|^{r+\eta-2(i_1+\cdots+i_n)} (y^1)^{i_1} \cdots (y^k)^{i_n}\,,$$

where

$$C^s_{i_1,\dots,i_n)^{s+1}} = \big(r+\eta-2(i_1+\cdots+i_k-1+\cdots+i_n)\big) C_{i_1,\dots,i_k-1,\dots,i_n} + (i_k+1) C^s_{i_1,\dots,i_k+1,\dots,i_n}\,.$$

From this formula, it follows first that there is a constant $K$ such that

$$|g(y)| \leq K |y|^{r-s+\eta}\,,$$

and then that $\tilde{f}_\eta$ has continuous partial derivatives of order up to and including $r$ on the range of the chart $\psi$.

[3] It seems clear that the subspace $\mathfrak{A}_1$ depends on the choice of the chart $\psi$.

We now know that $f_\eta$ is a $C^r$ function on the domain of the chart $\psi$.

We shall now prove that $f_\eta$ does not belong to $\mathfrak{A}_1$. It suffices to show that the pure derivative of $f_\eta$ of order $r+1$ on the variable $x^1$ does not exist at $P$. We do that by considering the function $g(t)$ of a real variable $t$ given by

$$g(t) = \tilde{f}_\eta(t, 0, \ldots, 0) = |t|^{r+\eta}\,.$$

(If $g^{(r+1)}(0)$ does not exist, then by definition, $\frac{\partial^{r+1} f_\eta}{\partial (x^1)^{r+1}}$ does not exist at $P$.) It is easy to show by induction that for $t \neq 0$,

$$g^{(s)}(t) = (r+\eta)(r+\eta-1)\cdots(r+\eta-s+1)|t|^{r+\eta-s}(\operatorname{sgn} t)^s\,.$$

It follows that $g^{(s)}(t) = 0$ for $s \leq r$. But then if $g^{(r+1)}(0)$ existed, we would have,

$$g^{(r+1)}(0) = (r+\eta)(r-1+\eta)\cdots(1+\eta)\lim_{t\to 0}\frac{|t|^\eta}{t}(\operatorname{sgn} t)^r\,.$$

and this last limit does not exist since the function whose limit is being taken is not bounded on any deleted neighborhood of $t = 0$.

Hence, as asserted, $f_\eta$ is not in $\mathfrak{A}_1$. Since it is a $C^r$ function, it is linearly independent of $\mathfrak{A}_1$. It is obvious that the whole set of functions $f_\eta$, $0 < \eta < 1$, is a linearly independent subset of the algebra $\mathfrak{A}$ and has cardinality $\mathfrak{c}$.

Let $\mathfrak{A}_2$ be the vector subspace (*not* a subalgebra!) of $\mathfrak{A}$ generated by the functions $f_\eta$, and let $\mathfrak{A}_3$ be a further vector subspace of $\mathfrak{A}$ such that $\mathfrak{A}$ has the direct-sum representation $\mathfrak{A} = \mathfrak{A}_1 \oplus \mathfrak{A}_2 \oplus \mathfrak{A}_3$. Thus, every element $f$ of $\mathfrak{A}$ has a unique representation as

$$f = f_1 + f_2 + f_3$$

with $f_i \in \mathfrak{A}_i$, $i = 1, 2, 3$.

We define a linear functional $L_\eta$ by specifying that

$$L_\eta(f) = \begin{cases} c & \text{if } f_2 = cf_\eta \text{ for some constant } c\,; \\ 0 & \text{if } f_2 \text{ is not a scalar multiple of } f_\eta\,. \end{cases}$$

As elements of $V$, the linear functionals $L_\eta$ form a linearly independent set, and hence span a subspace of dimension $\mathfrak{c}$. It remains to be shown that for any two $C^r$ functions $f$ and $g$ we have

$$L_\eta(fg) = f(P)L_\eta(g) + g(P)L_\eta(f)\,.$$

We write

$$fg = -f(P)g(P) + \big(f - f(P)\big)\big(g - g(P)\big) + f(P)g + g(P)f\,.$$

The first term on the right belongs to $\mathfrak{A}_1$ since it is constant. The second term belongs to $\mathfrak{A}_1$ by Lemma 4.2. It then follows that

$$L_\eta(fg) = f(P)L_\eta(g) + g(P)L_\eta(f)\,.$$

In this way, we have constructed a linearly independent subset of $V$ having cardinality $\mathfrak{c}$. Since the entire vector space $V$ has cardinality $\mathfrak{c}$, this must be its dimension as a real vector space. □

That the algebraic dimension of the space of derivations at a point on a manifold of finite smoothness is uncountably infinite is actually not surprising. Any infinite-dimensional Banach space (complete normed vector space) has an *algebraic* dimension that is uncountable, as follows easily from the Baire category theorem. The *geometric* dimension, that is, the cardinality of a linearly independent set whose finite linear combinations can *approximate* any element with arbitrary precision, is

countable if the space is separable, and such is the case with most of the important ones. (See, for example, Problem 3.19 of Appendix 3.)

**2.3. $C^\infty$ and $C^\omega$ manifolds.** Finally, we consider the derivations of the space of functions that are defined on a neighborhood of a point $P$ on a $C^\infty$ manifold of dimension $n$, where the neighborhood may depend on the function, and have continuous derivatives of all orders on that neighborhood. We observe that we can easily find $n$ linearly independent derivations at a point $P$. All we have to do is fix a chart $\psi$ and let

$$L_k(f) = \frac{\partial f}{\partial x^k}\,.$$

If $\psi(Q) = \big(\psi^1(Q), \dots, \psi^n(Q)\big)$, it is obvious that

$$\frac{\partial \psi^j}{\partial x^k} = \delta^j_k\,,$$

c since $\widetilde{\psi^j}(\boldsymbol{y}) = y^j$.

It follows that if $L(f) = c_1 \frac{\partial f}{\partial x^1} + \cdots + c_n \frac{\partial f}{\partial x^n} = 0$ for all $f$, then $c_j = L(\psi^j) = 0$ for all $j$. This much is true even on a $C^1$ manifold.

**Theorem 4.4.** *The space of derivations of $C^\infty$ functions at the point $P$ has dimension $n$, equal to that of the manifold and is spanned by the partial derivative operators $\frac{\partial}{\partial x^i}$.*

Proof. We need to show only that these derivations span the whole space of derivations, that is, that every derivation is a linear combination of them.

Let $f$ be any locally $C^\infty$ function on a neighborhood of $P$, which without loss of generality can be assumed to be the inverse image of a convex open set in $\mathbb{R}^n$ containing $\psi(P)$ and contained in the range of $\psi$. Fix a point $Q$ in this neighborhood, and let

$$\varphi(t) = \tilde{f}\big((1-t)\psi(P) + t\psi(Q)\big)\,.$$

We have

$$\begin{aligned} f(Q) &= f(P) + \varphi(1) - \varphi(0) \\ &= f(P) + \int_0^1 \varphi'(s)\,ds \\ &= f(P) + \sum_{k=1}^n \left( \int_0^1 \frac{\partial \tilde{f}}{\partial x^k}\big((1-s)\psi(P) + s\psi(Q)\big)\,ds \right) \big(\psi^k(Q) - \psi^k(P)\big) \\ &= f(P) + \sum_{k=1}^n \varphi_k(Q)\big(\psi^k(Q) - \psi^k(P)\big)\,. \end{aligned}$$

Here

$$\varphi_k(Q) = \int_0^1 \frac{\partial \tilde{f}}{\partial x^k}\big((1-s)\psi(P) + s\psi(Q)\big)\,ds\,,$$

so that, in particular, $\varphi_k(Q)$ is a $C^\infty$ function of $Q$, and

$$\varphi_k(P) = \frac{\partial \tilde{f}}{\partial x^k}\big(\psi(P)\big)\,.$$

Since $L(f(P)) = 0$ (because $f(P)$ is constant as a function of $Q$), the fact that $L$ is a derivation means that

$$L(f) = \sum_{k=1}^{n} \varphi_k(P) L(\psi^k) = \sum_{k=1}^{n} c_k \frac{\partial \tilde{f}}{\partial x^k}\big(\psi(P)\big)\,,$$

where $c_k = L(\psi^k)$. It follows that the space of derivations is spanned by the $n$ partial derivatives that were proved above to be linearly independent. Thus the dimension of the space of derivations (henceforth to be called the *tangent space at* $P$) is $n$, the same as the dimension of the manifold. □

Since a $C^\omega$ (analytic) manifold is in particular a $C^\infty$ manifold, and since the partial derivatives of analytic functions are also analytic, these partial derivatives also provide a basis for the space of derivations on a $C^\omega$ manifold.

**2.4. The tangent space to a manifold.** The set of derivations at a point $P$ on a $C^\infty$ manifold—the only manifolds we shall consider from now on—is a real vector space defined using properties intrinsic to the manifold itself and independent of any particular choice of a local chart at the point $P$. It can be pictured as the set of possible velocities at the point $P$ of a particle moving over the manifold. We call it the *tangent space* at $P$ since the velocity of a particle moving over a surface $S$ in $\mathbb{R}^3$ is represented by a vector tangent to $S$. If we were working with a manifold of finite smoothness, it would be necessary to *define* the tangent space at $P$ to be the subspace of the space of derivations spanned by the partial derivative operators. That is one reason why we don't work with these spaces. Another reason is that a partial derivative operator maps $C^r$ functions into $C^{r-1}$ functions and hence is not an operator on any reasonably defined vector space of functions.[4]

The connection with the geometric notion of a tangent space is an important one in making the transition from surfaces in $\mathbb{R}^3$ to abstract manifolds. Surfaces in $\mathbb{R}^3$ provide the justification for the more abstract notion of a derivation that we have just introduced. The connection is as follows: Given a surface with the parametrization $(u, v) \mapsto \boldsymbol{r}(u, v) = x(u, v)\boldsymbol{i} + y(u, v)\boldsymbol{j} + z(u, v)\boldsymbol{k} = \boldsymbol{r}(u, v)$, we naturally identify the derivation $X = a\partial/\partial u + b\partial/\partial v$ with the tangent vector to the surface $\boldsymbol{X} = a\partial\boldsymbol{r}/\partial u + b\partial\boldsymbol{r}/\partial v$. Formally, this identification means that for any differentiable function $f(u, v)$,

$$X(f) = \boldsymbol{X} \cdot \nabla \tilde{f}\,,$$

where $\tilde{f}(x, y, z)$ is any smooth function such that $\tilde{f}\big(\boldsymbol{r}(u, v)\big) = f(u, v)$ and $\nabla$ is the gradient operator:

$$\nabla \tilde{f} = \frac{\partial \tilde{f}}{\partial x}\boldsymbol{i} + \frac{\partial \tilde{f}}{\partial y}\boldsymbol{j} + \frac{\partial \tilde{f}}{\partial z}\boldsymbol{k}\,.$$

(This is Problem 4.8 below.)

[4] Newns and Walker made this definition seem more natural by defining the *cotangent* space at $P$ to be the quotient vector space of the space of locally $C^r$ functions at $P$ modulo the subspace of *stationary* functions (those having all first-order partial derivatives equal to zero at $P$). This quotient space is easily seen to be a vector space of dimension equal to the dimension of the manifold, and hence the tangent space, defined as the dual of the cotangent space, is also of that dimension. These definitions also provide an intrinsic definition of the tangent space. A derivation (such as $L_\eta$) that is not a linear combination of partial derivatives was called a *hypertangent vector* by Newns and Walker.

As a result, when the manifold we are discussing happens to be a parameterized surface in $\mathbb{R}^3$, we can use either the algebraic language of derivations, in which a contravariant vector $X$ has standard coordinates $(a, b)$, or the geometric language of vectors in $\mathbb{R}^3$, in which the same vector has components

$$\left(a\frac{\partial x}{\partial u} + b\frac{\partial x}{\partial v}, a\frac{\partial y}{\partial u} + b\frac{\partial y}{\partial v}, a\frac{\partial z}{\partial u} + b\frac{\partial z}{\partial v}\right).$$

The algebraic work involved in computing covariant derivatives and geodesics is greatly simplified by using the former notation, and that is what we shall generally do.

Regarded as an abstract vector space, the tangent space to an $n$-dimensional manifold at $P$ is simply the ordinary Euclidean space $\mathbb{R}^n$. As such, it has a dual space consisting of the set of all linear functions mapping it into the real numbers. The reader undoubtedly knows that this dual space has a natural embedding in $\mathbb{R}^n$ given by the dot product: If $\lambda : \mathbb{R}^n \to \mathbb{R}$ is any linear functional, there exists a unique vector $\boldsymbol{v}_\lambda \in \mathbb{R}^n$ such that

$$\lambda(\boldsymbol{u}) = \boldsymbol{v}_\lambda \cdot \boldsymbol{u}$$

for all $\boldsymbol{u} \in \mathbb{R}^n$.

In the light of this identification of $\mathbb{R}^n$ with its dual space, it might seem that we could ignore the distinction between vector fields and covector fields. It must be kept in mind, however, as noted in the quotation above from Weierstrass, that we cannot discuss any particular manifold in this abstract way. We have to introduce some parametrization or other in order to carry on a conversation about it. Any set of parameters $(x^1, \ldots x^n)$ provides a canonical basis of the tangent space, consisting of the operators

$$\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}.$$

If someone else is using a different set of parameters $(y^1, \ldots, y^n)$, then the canonical basis will be

$$\frac{\partial}{\partial y^1}, \ldots, \frac{\partial}{\partial y^n},$$

and there will be a natural mapping of parameters $(x^1, \ldots, x^n) \mapsto (y^1, \ldots, y^n) = \big(y^1(x^1, \ldots, x^n), \ldots, y^n(x^1, \ldots, x^n)\big)$ reflecting the conversion from one to the other.

The increment in the mapping $(x^1, \ldots, x^n) \mapsto (y^1, \ldots, y^n)$ near a fixed base point has a best linear approximation at each point whose matrix is the Jacobian matrix

$$\frac{\partial(y^1, \ldots, y^n)}{\partial(x^1, \ldots, x^n)} = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \cdots & \frac{\partial y^1}{\partial x^n} \\ \vdots & & \vdots \\ \frac{\partial y^n}{\partial x^1} & \cdots & \frac{\partial y^n}{\partial x^n} \end{pmatrix}.$$

The inverse mapping $(y^1, \ldots, y^n) \mapsto (x^1, \ldots, x^n) = \big(x^1(y^1, \ldots, y^n), \ldots, x^n(y^1, \ldots, y^n)\big)$ likewise has a best-linear approximation to its increment that is the inverse Jacobian:

$$\left(\frac{\partial(y^1, \ldots, y^n)}{\partial(x^1, \ldots, x^n)}\right)^{-1} = \frac{\partial(x^1, \ldots, x^n)}{\partial(y^1, \ldots, y^n)} = \begin{pmatrix} \frac{\partial x^1}{\partial y^1} & \cdots & \frac{\partial x^1}{\partial y^n} \\ \vdots & & \vdots \\ \frac{\partial x^n}{\partial y^1} & \cdots & \frac{\partial x^n}{\partial y^n} \end{pmatrix}.$$

**2.5. Reprise of covariance and contravariance.** Although we have discussed these concepts in several other places, it may be in order to say a few additional words at this point. By the chain rule,

$$\frac{\partial}{\partial y^i} = \frac{\partial x^1}{\partial y^i}\frac{\partial}{\partial x^1} + \cdots + \frac{\partial x^n}{\partial y^i}\frac{\partial}{\partial x^n}\,.$$

Thus, the $i$th canonical basis vector in $y$-coordinates is obtained by multiplying the canonical basis vectors in $x$-coordinates by the corresponding elements of the $i$th column of the *inverse* Jacobian matrix, the one that approximates the increment in the mapping from $y$ to $x$, which goes in the opposite direction. The mapping goes from $y$ to $x$, but the columns of its Jacobian map the basis vectors in $x$ coordinates to the basis vectors in $y$ coordinates. Because of that "oppositeness," we say that the tangent space is made up of *contravariant* vectors.

At this point, it will be well to head off the inevitable confusion that can arise when one tries to keep the two kinds of variance straight. It is very tempting to think of a tangent vector in $x$-parameters as $\boldsymbol{u}_x = u_x^1\partial/\partial x^1 + \cdots + u_x^n\partial/\partial x^n$. To do so is to create a linear mapping $\boldsymbol{u}_x \mapsto (u_x^1, \ldots, u_x^n) \in \mathbb{R}^n$. The problem with doing so is that if we now change coordinates and we wish to talk about the same derivation on $C^\infty$ in $y$-coordinates, we need to have, for any $C^\infty$ function $f$,

$$\begin{aligned}
\boldsymbol{u}_y f = \boldsymbol{u}_x f &= u_x^1\frac{\partial f}{\partial x^1} + \cdots + u_x^n\frac{\partial f}{\partial x^n} \\
&= \sum_{j=1}^n u_x^j\frac{\partial f}{\partial x^j} \\
&= \sum_{j=1}^n u_x^j\sum_{i=1}^n\frac{\partial y^i}{\partial x^j}\frac{\partial f}{\partial y^i} \\
&= \sum_{i=1}^n\left(\sum_{j=1}^n u_x^j\frac{\partial y^i}{\partial x^j}\right)\frac{\partial f}{\partial y^i}\,.
\end{aligned}$$

But since

$$\boldsymbol{u}_y f = \sum_{i=1}^n u_y^i\frac{\partial f}{\partial y^i}\,,$$

it follows that

$$u_y^i = \frac{\partial y^i}{\partial x^1}u_x^1 + \cdots + \frac{\partial y^i}{\partial x^n}u_x^n\,.$$

In other words, the $i$th *coefficient* of the vector in $y$-parameters is obtained by multiplying its coefficients in $x$-parameters by the elements in the corresponding *row* of the *direct* Jacobian. The coefficients of $\boldsymbol{u}$ are therefore called *covariant* functions of $\boldsymbol{u}$. It is actually the *mapping* $\boldsymbol{u} \to (u_x^1, \ldots, u_x^n)$ that is covariant, but as a matter of shorthand, we refer to the image under this mapping as a covariant function of the elements in the domain.

The lesson to be learned is that the coefficients of a *contravariant* vector transform *covariantly*. That is because they are covariant functions of the vectors whose components they are. We hope not to have to refer to this awkward fact very often, but it is necessary to mention it at least once. Some effort of will and patience is required to keep in mind that, when $u_x^i$ is the $i$th component of a *contra*variant vector field $\boldsymbol{u}$ in some parametrization, the *mapping* $\boldsymbol{u} \mapsto u_x^i$ is a *co*vector field. This

particular covector field is denoted $dx^i$ because of the standard notation involving the multi-variable chain rule (see below).

Again, there is an unfortunate danger of confusion when we use this notation, because we have all been trained to think of $dx^i$ as an infinitesimal increment in $x^i$. And, in fact, we shall have need for that interpretation occasionally. It turns out to be quite compatible with the new use of this symbol that we are about to introduce. For us, the symbols $dx^1, \dots, dx^n$ are to be regarded as the basis of the dual space to the tangent space that is the dual of its canonical basis in $x$-coordinates, that is, $\partial/\partial x^1, \dots, \partial/\partial x^n$. All that we really mean by them is contained in the relation

$$dx^i\Big(\frac{\partial}{\partial x^j}\Big) = \delta^i_j = \begin{cases} 0 & \text{if } i \neq j\,, \\ 1 & \text{if } i = j\,. \end{cases}$$

If $\boldsymbol{v} = \boldsymbol{v}_\lambda$ is the covector field corresponding to a field of linear functionals $\lambda$ on the tangent space, the well-known relation

$$\boldsymbol{v} \cdot T\boldsymbol{u} = T^\tau \boldsymbol{v} \cdot \boldsymbol{u}\,,$$

where $T^\tau$ denotes the transpose of the operator $T$, shows that when we change coordinates, the covector $\boldsymbol{v} = a_1\, dx^1 + \cdots a_n\, dx^n$ transforms into $b_1\, dy^1 + \cdots b_n\, dy^n$, where

$$\begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \cdots & \frac{\partial y^1}{\partial x^n} \\ \vdots & & \vdots \\ \frac{\partial y^n}{\partial x^1} & \cdots & \frac{\partial y^n}{\partial x^n} \end{pmatrix}.$$

Therefore, we can keep our records straight by identifying a vector field with the column matrix of its components and a covector field as the row matrix of those components. When we do that, the effect of allowing the functional $\lambda$ corresponding to the covector $\boldsymbol{v}$ whose components are $a_1, \dots, a_n$ to operate on the vector $\boldsymbol{u}$ whose components are $c^1, \dots, c^n$ is just the matrix product

$$\begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} c^1 \\ \vdots \\ c^n \end{pmatrix} = a_1 c^1 + \cdots + a_n c^n = \boldsymbol{v} \cdot \boldsymbol{u}\,.$$

**Remark 4.5.** For a $k$-dimensional manifold defined by an embedding in $\mathbb{R}^n$, there are two ways of regarding the tangent space at $P$ as a vector space of dimension $k$. One, which is absolute as far as the manifold is concerned, is to use the embedding $(x^1, \dots, x^k) \mapsto \boldsymbol{r}(x^1, \dots, x^k)$ and take the tangent space to be the $k$-dimensional subspace of $\mathbb{R}^n$ spanned by the vectors

$$\frac{\partial \boldsymbol{r}}{\partial x^1}\,, \dots\,, \frac{\partial \boldsymbol{r}}{\partial x^k}\,,$$

all evaluated at the point $(x^1, \dots, x^k)$ such that $P = \boldsymbol{r}(x^1, \dots, x^k)$. Algebraically this procedure amounts to moving the origin of $\mathbb{R}^n$ to the point $P$. It is easy to visualize.

The other way is to think of the tangent vector as we have described it here: a derivation operator on $C^\infty$ functions defined on a neighborhood of $P$. The two are trivially equivalent under the correspondence

$$u^1 \frac{\partial \boldsymbol{r}}{\partial x^1} + \cdots + u^n \frac{\partial \boldsymbol{r}}{\partial x^k} \leftrightarrow u^1 \frac{\partial}{\partial x^1} + \cdots + u^n \frac{\partial}{\partial x^k}\,,$$

and either of these expressions can justifiably be denoted as a generic tangent vector or vector field $\boldsymbol{u}$. Either way that we do it, we need the parameters $x^1, \dots, x^n$. The expression as a derivation, however, eliminates the excess baggage that the mapping $\boldsymbol{r}$ carries with it in the form of extra coordinates in $\mathbb{R}^n$. If all the information we need can be articulated without that extra baggage, it is desirable to do so. That is the chief aim of Chapter 5 in Volume 1, where we labor to express the curvature of a surface in $\mathbb{R}^3$ without having to make use of the third dimension.

As long as we are trying to establish an intuitive basis for our procedures using the model of two-dimensional surfaces in $\mathbb{R}^3$, we shall retain the extra coordinate. But it is desirable to throw away that crutch as soon as possible and use only the minimal information that comes from the parameters. We make that transition gradually in the course of Chapters 5 and 6 of Volume 1.

**2.6. The inner product.** We define the standard inner product of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^k$ to be

$$\boldsymbol{v}(\boldsymbol{u}) = \boldsymbol{u}(\boldsymbol{v}) = u^1 v_1 + \cdots + u^k v_k \,.$$

It is tempting to think of this interaction as the dot product on $\mathbb{R}^k$, with $\boldsymbol{u}$ and $\boldsymbol{v}$ identified with the vectors $(u^1, \dots, u^k)$ and $(v_1, \dots, v_k)$. To do this is to use the identification of a linear functional (covector) $\lambda$ on $\mathbb{R}^k$ with the unique vector $(u^1_\lambda, \dots, u^k_\lambda) \in \mathbb{R}^k$ such that $\lambda(v^1, \dots, v^k) = u^1_\lambda v^1 + \cdots + u^k_\lambda v^k$. One way to make this identification appear more natural is to introduce the notion of lowering the contravariant indices of a vector using the metric tensor $g_{ij}$. We make the definition

$$u^*_j = g_{ij} u^i \,,$$

where, of course, summation on the index $i$ extends from 1 to $k$. In this way a contravariant vector $u^i \partial/\partial x^i$ gets "lowered" to create a covector $u^*_j \, dx^j$. In Euclidean space, where $g_{ij} = \delta^j_i$, the above-mentioned identification of the dual space with the original space amounts to just this lowering.

Being a covector, $\boldsymbol{u}^*$ can act on a vector $\boldsymbol{w} = w^j \, \partial/\partial x^j$ to yield the *first fundamental form*:

$$\boldsymbol{u}^*(\boldsymbol{w}) = g_{ij} u^i w^j = \langle \boldsymbol{u}, \boldsymbol{w} \rangle = g\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u} \cdot g\boldsymbol{v} \,,$$

where $g$ denotes the linear operator given by the equation

$$g(\boldsymbol{u}) = \begin{pmatrix} g_{11} & \cdots & g_{1k} \\ \vdots & & \vdots \\ g_{k1} & \cdots & g_{kk} \end{pmatrix} \begin{pmatrix} u^1 \\ \vdots \\ u^k \end{pmatrix} .$$

If the metric coefficients $g_{ij}$ are defined as the dot products of the basis vectors $(\partial \boldsymbol{r}/\partial x^i) \cdot (\partial \boldsymbol{r}/\partial x^j)$, then the operator $g$ is symmetric with respect to the dot product and indeed positive definite, since $g(\boldsymbol{u}) \cdot \boldsymbol{u} = \left| \sum_i u^i \partial \boldsymbol{r}/\partial x^i \right|^2$, which is positive unless $\boldsymbol{u} = \boldsymbol{0}$. Thus $\langle \cdot \,, \cdot \rangle$ is a positive-definite bilinear form on the tangent space, and it provides the natural definition of the length of a vector, a definition that is independent of the local parametrization of the manifold. In particular, when we consider a path $\boldsymbol{\gamma}(t)$ parameterized as $\big(x^1(t), \dots, x^k(t)\big)$, that length is what one would find for the derivative vector $\boldsymbol{\gamma}'(t)$ if $\boldsymbol{\gamma}(t)$ were regarded as a path in the ambient space $\mathbb{R}^n$ and identified with $\boldsymbol{r}\big(\boldsymbol{\gamma}(t)\big)$. We would have

$$|\boldsymbol{\gamma}'(t)|^2 = g_{ij}\big(x^1(t), \dots, x^k(t)\big)(x^i)'(t)(x^j)'(t) = \frac{d\big(\boldsymbol{r}(\boldsymbol{\gamma}(t))\big)}{dt} \cdot \frac{d\big(\boldsymbol{r}(\boldsymbol{\gamma}(t))\big)}{dt} \,,$$

which is the inner product defined by use of $g = (g_{ij})$ that we shall rely on when discussing the geometry of tangent fields.

## 3. Tensors

In multilinear algebra, which is an essential tool for understanding differential geometry and relativity, the relationship between a finite-dimensional real vector space $X$ and its dual space $X^*$ is fundamental. The space $X^*$ consists of linear functionals on $X$, that is, linear transformations from $X$ to $\mathbb{R}$. The connection with coordinate systems arises when two observers reconcile their bookkeeping. If the space of measurements made by one observer is $X$, and those made by a second observer form the vector space $Y$, and the reconciliation is effected, as if often is, through a linear transformation $T : X \to Y$, there is a natural linear transformation $T^*$ from $Y^*$ to $X^*$, called the *adjoint* of $T$ and defined by

$$T^*(L)(\boldsymbol{x}) = L\big(T(\boldsymbol{x})\big)\,.$$

It is easy to verify that $T^*(L)$ belongs to $X^*$ for all $L \in Y^*$. Since $T^*$ maps in the opposite direction from $T$, it is said to be *contravariant*.

Without bothering to explore this topic in its greatest generality, we are going to confine ourselves to what is essential when coordinates are changed. Thus, we consider two sets of coordinates $\boldsymbol{x} = (x^1, \ldots, x^n)$ and $\boldsymbol{y} = (y^1, \ldots, y^n)$ on an open subset of a $C^\infty$ manifold. By the chain rule,

$$\frac{\partial}{\partial y^j} = \sum_{i=1}^{n} \frac{\partial x^i}{\partial y^j} \frac{\partial}{\partial x^i}\,.$$

It follows that the gradient operators in the two coordinate systems are related by

$$\begin{aligned}
\nabla_{\boldsymbol{y}} &= \begin{pmatrix} \frac{\partial}{\partial y^1} & \cdots & \frac{\partial}{\partial y^n} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial}{\partial x^1} & \cdots & \frac{\partial}{\partial x^n} \end{pmatrix} \begin{pmatrix} \frac{\partial x^1}{\partial y^1} & \cdots & \frac{\partial x^1}{\partial y^n} \\ \vdots & & \vdots \\ \frac{\partial x^n}{\partial y^1} & \cdots & \frac{\partial x^n}{\partial y^n} \end{pmatrix} \\
&= \nabla_{\boldsymbol{x}} \frac{\partial(x^1, \ldots, x^n)}{\partial(y^1, \ldots, y^n)}\,.
\end{aligned}$$

In contrast to the partial derivative operator, the differential $dy^i$ transforms (again via the chain rule) as

$$dy^i = \sum_{j=1}^{n} \frac{\partial y^i}{\partial x^j}\, dx^j\,.$$

In matrix form, we have

$$d\boldsymbol{y} = \begin{pmatrix} dy^1 \\ \vdots \\ dy^n \end{pmatrix} = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \cdots & \frac{\partial y^1}{\partial x^n} \\ \vdots & & \vdots \\ \frac{\partial y^n}{\partial x^1} & \cdots & \frac{\partial y^n}{\partial x^n} \end{pmatrix} \begin{pmatrix} dx^1 \\ \vdots \\ dx^n \end{pmatrix} = \frac{\partial(y^1, \ldots, y^n)}{\partial(x^1, \ldots, x^n)} d\boldsymbol{x}\,.$$

For that reason, we say that a partial derivative is a contravariant vector, while a differential is a covariant vector. As we have discussed the contrast between these

two kinds of vectors in several different places, we say no more about it at this point.

## 4. Parallel Transport

Physicists and mathematicians are accustomed to the use of what are called *free* vectors in $\mathbb{R}^3$. Two vectors representing forces applied at different points of a body, for example, can be regarded as equal if they have the same length and direction, which is to say, $\boldsymbol{v}_1$ can be transported from a point $P_1 \in \mathbb{R}^3$ to $P_2 \in \mathbb{R}^3$ without changing its length or its orientation in $\mathbb{R}^3$, and will coincide with $\boldsymbol{v}_2$ after the transport. In general, as noted above, there is not any parameter-independent sense in which a tangent vector $\boldsymbol{v}_1$ at a point $P_1$ on a manifold is the "same vector" as a tangent vector $\boldsymbol{v}_2$ at a point $P_2$. Nevertheless, in Chapter 6 of Volume 1, we introduced the notion of parallel transport and showed how the covariant derivative allows us to make sense of the concept of moving a contravariant vector from the tangent space at one point of a manifold to the tangent space at another point without changing the vector. For ease of reference, here is the corresponding definition:

**Definition 4.1.** Let $U$ be a connected open set in $\mathbb{R}^n$ that parameterizes a portion of a manifold $\mathfrak{M}$ containing points $P_1$ and $P_2$. If $\boldsymbol{v}_1$ is a tangent vector at $P_1$ and $\boldsymbol{v}_2$ is a tangent vector at $P_2$, and $\boldsymbol{\gamma} : [a, b] \to U$ is a curve such that $\boldsymbol{v}\big(\boldsymbol{\gamma}(a)\big) = \boldsymbol{v}_1$, $\boldsymbol{v}\big(\boldsymbol{\gamma}(b)\big) = \boldsymbol{v}_2$, and

$$(\gamma^j)'(t)\Big(\frac{\partial v^k(\boldsymbol{\gamma}(t))}{\partial x^j} + v^i(\boldsymbol{\gamma}(t))\Gamma^k_{ij}\big(\boldsymbol{\gamma}(t)\big)\Big) \equiv 0$$

for all $t \in (a, b)$ and all $k = 1, \dots, n$. (Here the Einstein summation convention is in effects and the summations on $i$ and $j$ range from 1 to $n$.)

As an easy mnemonic, we write this condition as $\nabla_{\boldsymbol{\gamma}'(t)}\boldsymbol{v}\big(\boldsymbol{\gamma}(t)\big) \equiv 0$, and this equation is literally true if $\boldsymbol{v}$ is a smooth vector field defined in a neighborhood of the curve traced by $\boldsymbol{\gamma}(t)$. But, in general, the condition in the definition amounts to a set of differential equations that, together with the initial condition $\boldsymbol{v}\big(\boldsymbol{\gamma}(a)\big) = \boldsymbol{v}_1$ define $\boldsymbol{v}\big(\boldsymbol{\gamma}(t)\big)$ for all $t$, without defining $\boldsymbol{v}$ at any point not on the curve. To describe this concept intuitively, we say that the projection of the derivative on the tangent vector to the curve is constant along the curve. That this definition is parameter-independent follows from the parameter-independence of the covariant derivative. To use informal language, the covariant derivative of $\boldsymbol{v}\big(\boldsymbol{\gamma}(t)\big)$ vanishes along the curve, and hence the vector field is constant along the curve (as far as the denizens of the surface under consideration can tell), so that it is natural to regard $\boldsymbol{v}(a) = \boldsymbol{v}_1$ and $\boldsymbol{v}(b) = \boldsymbol{v}_2$ as the same vector.

More generally, the covariant derivative along the curve is a measure of the extent to which length and direction vary as the curve is traversed. As one can easily see, under parallel transport, the length will not change at all. In that case, parallel transport around a closed loop will reveal the extent to which direction has altered, in a way insensible to those living on the manifold, during the traversal. If there is a two-dimensional submanifold (surface) that contains the loop, this alteration is a measure of the curvature of that surface, as shown in Theorem 6.6 of Volume 1. Illustrations of this principle are given in Examples 4.6–4.9 below.

The parallel transport of a vector is parameter-independent, which is a crucial fact when it is necessary to transport a vector over a curve that is not contained

in the domain of any one local parametrization of a manifold. It will be necessary at some point to switch parameters, and so it is important to know that on the portion of the curve that lie in the image of two different parametrizations, both parametrizations yield the same tangent vectors.

If, in some local set of parameters $(x^1, \dots, x^n)$ the vector field $\boldsymbol{v}$ is given by

$$\boldsymbol{v}(x^1, \dots, x^n) = v^i(x^1, \dots, x^n)\frac{\partial}{\partial x^i}$$

then the system of equations for parallel transport along $\boldsymbol{\gamma}(t) = \big(x^1(t), \dots, x^n(t)\big)$ says

$$(x^j)'(t)\Big(\frac{\partial w^i}{\partial x^j} + w^k\Gamma^i_{jk}\Big) = 0\,.$$

If we introduce the notation $\boldsymbol{w}(t) = \boldsymbol{w}\big(\boldsymbol{\gamma}(t)\big)$, the first term in this expression is simply $dw^i/dt$, and so this system of equations really says

$$\frac{dw^i}{dt} + (x^j)'(t)w^k(t)\Gamma^i_{jk}\big(\boldsymbol{\gamma}(t)\big) = 0\,. \tag{2}$$

As a consequence, the system of equations for parallel transport makes no reference to any points except those on the curve itself, and $\boldsymbol{w}$ need not be defined on an open set containing the curve. As a matter of fact, this system of equations together with the initial condition $\boldsymbol{w}(0) = \boldsymbol{w}_0$ *creates* the parallel transport of any vector along any curve. Moreover, since it is a linear system of equations, it follows that parallel transport over $\boldsymbol{\gamma}$ is a linear mapping and that there are positive constants $C$ and $K$, such that for any two vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ that are parallel-transported over $\boldsymbol{\gamma}$, the following inequality holds for all parameter values $t$ satisfying $|t| \le C$:

$$|\boldsymbol{w}_1(t) - \boldsymbol{w}_2(t)| \le K|\boldsymbol{w}_1(0) - \boldsymbol{w}_2(0)|\,.$$

(See the existence and uniqueness theorem for initial-value problems in Appendix 5 and the discussion of linear equations that follows it.)

**Example 4.6.** As the simplest non-trivial example, consider the sphere of radius $R$ about $(0,0,0)$ in $\mathbb{R}^3$ with the parametrization

$$(s,t) \mapsto \boldsymbol{r}(s,t) = (R\cos s\,\cos t, R\cos s\,\sin t, R\sin s)\,,$$

so that the standard ordered basis of the tangent space at $P = \boldsymbol{r}(s,t)$ is

$$\begin{aligned}\frac{\partial \boldsymbol{r}}{\partial s} &= (-R\sin s\,\cos t, -R\sin s\sin t, R\cos s)\,,\\ \frac{\partial \boldsymbol{r}}{\partial t} &= (-R\cos s\,\sin t, R\cos s\,\cos t, 0)\,.\end{aligned}$$

We observe that the vector $\partial\boldsymbol{r}/\partial s$ is of constant length, and therefore both of its partial derivatives are perpendicular to it. As a result, it follows that $\Gamma^1_{11} = 0 = \Gamma^1_{12} = \Gamma^1_{21}$. As it happens, $\partial^2\boldsymbol{r}/\partial s^2 = -\boldsymbol{r}$, which is perpendicular to $\partial\boldsymbol{r}/\partial t$, and so $\Gamma^2_{11} = 0$. Likewise, for each fixed value of $s$, $\partial\boldsymbol{r}/\partial t$ is of constant length, and hence its partial derivative with respect to $t$ is perpendicular to it. Therefore $\Gamma^2_{22} = 0$. That leaves only three non-zero Christoffel symbols, which are easy to compute: $\Gamma^2_{12} = \Gamma^2_{21} = -\tan s$, $\Gamma^1_{22} = \cos s\,\sin s$

Consider a fixed vector $\boldsymbol{v} = a\,\partial\boldsymbol{r}/\partial s + b\,\partial\boldsymbol{r}/\partial t$. Let us transport this vector along a line of longitude $t = \theta$ using the latitude $s$ as parameter. It is convenient to "lift" the curve $\boldsymbol{\gamma}(s)$ to the parameter space, where it is simply $\boldsymbol{\gamma}(s) = (s,\theta)$. Then

$\boldsymbol{\gamma}'(s) = (1,0)$, and—letting $\Pi$ denote the orthogonal projection onto the tangent space at a point—we find that the vector field $\boldsymbol{v}\big(\boldsymbol{\gamma}(s)\big) = a(s)\,\partial/\partial s + b(s)\,\partial/\partial t$ that is a parallel transport of this vector causes the following vector field to vanish identically:

$$\begin{aligned}\nabla_{\boldsymbol{\gamma}'(s)} &= \Pi\Big(\frac{\partial}{\partial s}\Big(a(s)\frac{\partial}{\partial s} + b(s)\frac{\partial}{\partial t}\Big)\\ &= \big(a'(s) + a(s)\Gamma^1_{11} + b(s)\Gamma^1_{12}\big)\frac{\partial}{\partial s} + \big(b'(s) + a(s)\Gamma^2_{11} + b(s)\Gamma^2_{12}\big)\frac{\partial}{\partial t}\\ &= a'(s)\frac{\partial}{\partial s} + \big(b'(s) - b(s)\,\tan s\big)\frac{\partial}{\partial t}\,.\end{aligned}$$

We conclude that $a(s)$ is constant and equal to $a$ at every point, while $b(s) = C\,\sec(s)$ for some constant $C$. At the parameter value $s_0$ where $b(s_0) = b$ (the initial point), we have $C = b\,\cos(s_0)$, and so $b(s) = b\cos(s_0)/\cos(s)$.

**Remark 4.6.** It is clear from this last expression that $b(s)$ is not defined at the North Pole where $s = \pi/2$, and in any case, no value of $t$ corresponds to this point. (This kind of parameter breakdown is precisely why manifolds are preferable to parametrizations in theoretical work.) However, the vector

$$C\,\sec(s)\frac{\partial \boldsymbol{r}}{\partial t} = CR(-\sin\theta, \cos\theta, 0)$$

has a constant value at every point along the line of longitude $t = \theta$ except at the North Pole, and therefore we are justified in using this constant value as the "end-value" for parallel transport along that line of longitude.

If the vector is transported to the North Pole parallel to itself, it becomes

$$\begin{aligned}a(-R\cos\theta, -R\sin\theta, 0) + bR(-\cos(s_0)\sin\theta, \cos(s_0)\cos\theta, 0) =\\ - R\big(a\cos\theta + b\cos(s_0)\sin\theta, a\sin\theta - b\cos(s_0)\cos\theta, 0\big)\,.\end{aligned}$$

Note also that the length of the transported vector, regarded as a vector in $\mathbb{R}^3$, is $R\sqrt{a^2 + b^2\cos^2 s_0}$ at every point.

**Example 4.7.** Let us calculate the result of a parallel transport of the vector $\boldsymbol{v} = a(\partial \boldsymbol{r}/\partial s) + b(\partial \boldsymbol{r}/\partial t)$ along the equator whose equation is $s = 0$ between two lines of longitude $t = \theta_1$ and $t = \theta_2$. Using $t$ as a parameter and taking $s = 0$, we have (again, moving $\boldsymbol{\gamma}$ into the parameter space), $\boldsymbol{\gamma}(t) = (0,t)$, so that $\boldsymbol{\gamma}'(t) = (0,1)$. Thus, we find that

$$\begin{aligned}\nabla_{\boldsymbol{\gamma}'(t)}\Big(a(t)\frac{\partial}{\partial s} + b(t)\frac{\partial}{\partial t}\Big) &= \Pi\Big(\frac{\partial}{\partial t}\Big(a(t)\frac{\partial}{\partial s} + b(t)\frac{\partial}{\partial t}\Big)\Big)\\ &= \big(a'(t) + a(t)\Gamma^1_{12} + b(t)\Gamma^1_{22}\big)\frac{\partial}{\partial s} + \big(b'(t) + a(t)\Gamma^2_{12} + b(t)\Gamma^2_{22}\big)\frac{\partial}{\partial t}\,.\end{aligned}$$

Since $\Gamma^1_{22} = \cos s\,\sin s = 0$ and $\Gamma^2_{12} = -\tan s = 0$ when $s = 0$, we therefore have the trivial differential equations $a'(t) = 0$, $b'(t) = 0$. That is, $a(t)$ and $b(t)$ remain constant. Notice that the vector we are translating, regarded as a vector in $\mathbb{R}^3$, has changed from

$$(0,0,aR) + (-bR\sin\theta_1, bR\cos\theta_1, 0) = R(-b\sin\theta_1, b\cos\theta_1, a)$$

to $R(-b\sin\theta_2, b\cos\theta_2, a)$.

The length of the vector therefore does not change, remaining equal to $R\sqrt{a^2+b^2}$ throughout. When $a$ is zero, we are parallel-translating the tangent vector to the equator, and the angle between the initial and final positions is $|\theta_1 - \theta_2|$, as one can easily tell by taking the dot product of the two vectors.

**Remark 4.7.** If we transport the vector $a\,\partial/\partial s + b\,\partial/\partial t$ from the point at latitude 0 and longitude $\theta_1$ straight north to the North Pole, it winds up, when regarded as a vector in $\mathbb{R}^3$, as the vector $-R\big(a\cos\theta_1 + b\sin\theta_1, a\sin\theta_1 - b\cos\theta_1, 0\big)$, but if we first transport it along the equator to longitude $\theta_2$ and then straight north to the north pole, it winds up as $-R\big(a\cos\theta_2 + b\sin\theta_2, a\sin\theta_1 - b\cos\theta_1, 0\big)$. The cosine of the angle $\varphi$ between these two vectors is their dot product divided by the product of their lengths, that is,

$$\frac{R^2\big((a\cos\theta_1 + b\sin\theta_1)(a\cos\theta_2 + b\sin\theta_2) + (a\sin\theta_1 - b\cos\theta_1)(a\sin\theta_2 - b\cos\theta_2)\big)}{R^2(a^2+b^2)} = \cos(\theta_2 - \theta_1)\,,$$

which is to say, the angle $\varphi$ between them is precisely the absolute value of the difference in the two longitudes.

Since the sum of the angles of this right spherical triangle is $\pi + |\theta_2 - \theta_1|$, this difference in longitudes is the angular excess of the triangle, which, as we know, is directly proportional to its area. That fact holds in general.

**Example 4.8.** As another simple example, consider a cylinder of radius $R$ in $\mathbb{R}^3$, whose parametrization is $\boldsymbol{r}(s,t) = (R\cos s, R\sin s, at)$, $-\infty < s < \infty$, $-\infty < t < \infty$. (This parametrization actually covers the cylinder infinitely many times, but it is locally one-to-one.) As this is a surface of curvature zero, we may expect some things to simplify considerably, and they do. To begin with, the standard ordered basis of tangent vectors is very simple, the two vectors being mutually orthogonal:

$$\begin{aligned}\frac{\partial \boldsymbol{r}}{\partial s} &= (-R\sin s, R\cos s, 0)\,,\\ \frac{\partial \boldsymbol{r}}{\partial t} &= (0,0,a)\,.\end{aligned}$$

We notice immediately that

$$\frac{\partial^2 \boldsymbol{r}}{\partial s^2} = (-R\cos s, -R\sin s, 0) = \frac{1}{a}\,\frac{\partial \boldsymbol{r}}{\partial t} \times \frac{\partial \boldsymbol{r}}{\partial s}\,,$$

and all other second-order partial derivatives of $\boldsymbol{r}$ vanish. Thus all second-order partial derivatives of $\boldsymbol{r}$ are perpendicular to the surface, and their projections into the tangent plane are zero. It follows that all the Christoffel symbols for this surface vanish, making the covariant derivatives very easy to take.

Let us imitate what we did above on the sphere, parallel-translating the vector $(1/R, 1/a) = (0,1,1)$ from $P_1 = \boldsymbol{r}(0,0) = (R,0,0)$ over the circle $t = 0$, $0 \le s \le \pi/2$ to the point $P_2 = \boldsymbol{r}(\pi/2, 0) = (0,R,0)$, then from $P_2$ to $P_3 = \boldsymbol{r}(\pi/2, 1) = (0,R,a)$ along the vertical line $s = \pi/2$, $0 \le t \le 1$, and then directly from $P_1$ to $P_3$ along the (geodesic) helix whose parametrization is $\boldsymbol{\gamma}(s) = (R\cos s, R\sin s, 2as/\pi)$, $0 \le s \le \pi/2$. We assume a vector field

$$\boldsymbol{v}(s,t) = a(s,t)\,\frac{\partial \boldsymbol{r}}{\partial s} + b(s,t)\,\frac{\partial \boldsymbol{r}}{\partial t}\,.$$

Along the path from $P_1$ to $P_2$, we take $\boldsymbol{\gamma}(s) = \boldsymbol{r}(s,0) = (R\cos s, R\sin s, 0)$, so that

$$\boldsymbol{\gamma}'(s) = (-R\sin s, R\cos s, 0) = \frac{\partial \boldsymbol{r}}{\partial s} = (1,0)$$

for each value of $s$. Then

$$\nabla_{(1,0)}\big(\boldsymbol{\gamma}(s)\big) = \tilde{a}'(s)\frac{\partial \boldsymbol{r}}{\partial s} + \tilde{b}'(s)\frac{\partial \boldsymbol{r}}{\partial t}\,,$$

where $\tilde{a}(s) = a(s,0)$ and $\tilde{b}(s) = b(s,0)$. Once again, trivially, $a(s)$ and $b(s)$ must be constant. Since we started with the vector $(1/R, 1/a)$, we have $a(s) \equiv 1/R$ and $b(s) \equiv 1/a$ for all $s$. Therefore the result of the parallel transport is the vector

$$\frac{1}{R}\,\frac{\partial \boldsymbol{r}}{\partial s}\Big|_{(s,t)=(\pi/2,0)} + \frac{1}{a}\,\frac{\partial \boldsymbol{r}}{\partial t}\Big|_{(s,t)=(\pi/2,0)} = (-1,0,0) + (0,0,1) = (-1,0,1)\,.$$

This is the same result we got on the sphere, which is no surprise, since the sphere and the cylinder we are considering coincide along this curve.

For the transport from $P_2$ to $P_3$, we have $\boldsymbol{\gamma}(t) = (0, R, at)$,

$$\boldsymbol{\gamma}'(t) = (0,0,a) = \frac{\partial \boldsymbol{r}}{\partial t}\,,$$

and once again, $\boldsymbol{\gamma}'(t)$ has coordinates $(0,1)$ at every value of $t$.

As before, if $\tilde{a}(t) = a(\pi/2, t)$ and $\tilde{b}(t) = b(\pi/2, t)$, then

$$\nabla_{(0,1)}\big(\boldsymbol{\gamma}(t)\big) = \tilde{a}'(t)\,\frac{\partial \boldsymbol{r}}{\partial s} + \tilde{b}'(t)\,\frac{\partial \boldsymbol{r}}{\partial t}\,,$$

and we conclude, once again that $a(t)$ and $b(t)$ have the constant values that they have at $t = 0$, which are respectively $1/R$ and $1/a$.

Thus the result of the parallel transport from $P_2$ to $P_3$ is

$$\frac{1}{R}\,\frac{\partial \boldsymbol{r}}{\partial s}\Big|_{(s,t)=(\pi/2,1)} + \frac{1}{a}\,\frac{\partial \boldsymbol{r}}{\partial t}\Big|_{(s,t)=(\pi/2,1)} = (-1,0,0) + (0,0,1) = (-1,0,1)\,.$$

In other words, the vector has not changed at all as a vector in $\mathbb{R}^3$. This should be no surprise, since the tangent planes to this cylinder at $P_2$ and $P_3$ and all points in between on this path are all one and the same.

For the third portion of this example, we take $\boldsymbol{\gamma}(s) = (R\cos s, R\sin s, 2as/\pi)$, so that

$$\boldsymbol{\gamma}'(s) = (-R\sin s, R\cos s, 2a/\pi) = \frac{\partial \boldsymbol{r}}{\partial s} + \frac{2}{\pi}\,\frac{\partial \boldsymbol{r}}{\partial t}\,.$$

Once again, we are fortunate enough to get a tangent vector whose coordinates are the same at each point. With the now-familiar notation, we have

$$\nabla_{(1,2/\pi)}\big(\boldsymbol{\gamma}(s)\big) = \tilde{a}'(s)\,\frac{\partial \boldsymbol{r}}{\partial s} + \frac{2}{\pi}\,\tilde{b}'(s)\,\frac{\partial \boldsymbol{r}}{\partial t}\,.$$

Thus, it is, yet again, no surprise that the coordinates of the parallel-transported vector remain constant. Since we are translating the vector $(0,1,1)$, whose coordinates are $1/R$, $1/a$, the result of the transport is

$$\frac{1}{R}\,\frac{\partial \boldsymbol{r}}{\partial s}\Big|_{(s,t)=(\pi/2,a)} + \frac{1}{a}\,\frac{\partial \boldsymbol{r}}{\partial t}\Big|_{(s,t)=(\pi/2,a)} = (-1,0,0) + (0,0,1) = (-1,0,1)\,.$$

This time, the result of the parallel transport is the same over both paths. Again, this should not be surprising, since the cylinder is a flat surface.

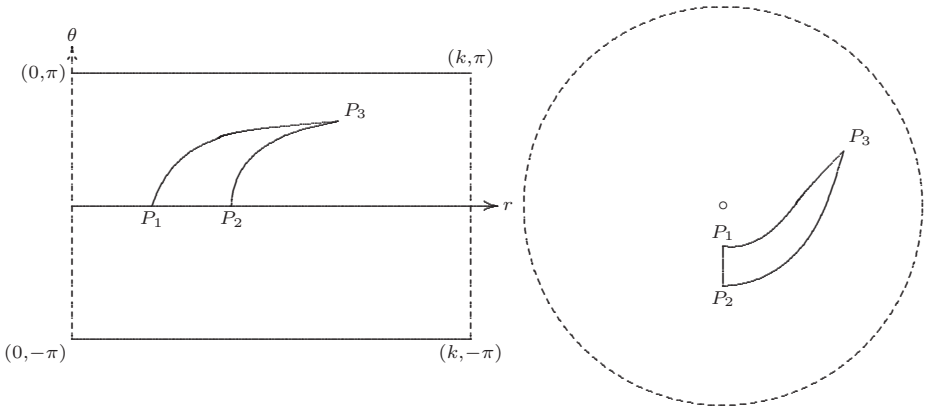


FIGURE 4. Left: The preimage in the space of parameters $r$, $\theta$ ($0 < r < k$, $-\pi \le \theta \le \pi$) of the triangle $P_1P_2P_3$ on the pseudo-hemisphere. Right: The projection of the triangle into the equatorial disk of the pseudo-hemisphere.

**Example 4.9.** Finally, to take one slightly less trivial example, we shall parallel-transport a vector around a geodesic right triangle on the pseudo-hemisphere, which is a model for a portion of the hyperbolic plane. If we use the parametrization

$$\boldsymbol{r}(r,\theta) = \left( r\cos\theta, r\sin\theta, k\left( \ln\left( \frac{k}{r} + \sqrt{\left(\frac{k}{r}\right)^2 - 1} \right) - \sqrt{1 - \left(\frac{r}{k}\right)^2} \right) \right),$$

that is, $r$ is the radial coordinate and $\theta$ the polar angle of the orthogonal projection of a typical point onto the equatorial disk of the pseudo-hemisphere, then the metric coefficients are $E(r,\theta) = k^2/r^2$, $F(r,\theta) = 0$, $G(r,\theta) = r^2$. The Christoffel symbols are as follows:

$$\Gamma^1_{11} = -\tfrac{1}{r}\,; \quad \Gamma^1_{12} = 0\,; \quad \Gamma^1_{21} = 0\,; \quad \Gamma^1_{22} = -\tfrac{r^3}{k^2}\,;$$
$$\Gamma^2_{11} = 0\,; \quad \Gamma^2_{12} = \tfrac{1}{r}\,; \quad \Gamma^2_{21} = \tfrac{1}{r}\,; \quad \Gamma^2_{22} = 0$$

The triangle we pick has vertices at $P_1 = \boldsymbol{r}(k/5, 0)$, $P_2 = \boldsymbol{r}(2k/5, 0)$, and $P_3 = \boldsymbol{r}(2k/3, 2)$. The angle at $P_2$ is a right angle.

As shown in Appendix 1, the geodesics on this surface are of two types: (1) lines of longitude, such as the side $P_1P_2$ of this triangle, for which the parameter $\theta$ is constant; (2) curves satisfying an equation of the form

$$r = \frac{kr_0}{\sqrt{k^2 - r_0^2(\theta - \theta_0)^2}}\,,$$

where $\theta_0$ is the value of $\theta$ for which $r$ is minimized, and $r_0$ is that minimum value of $r$.

It is easy to see that this curve is orthogonal to the line of longitude $\theta = \theta_0$. From the formulas given in Appendix 1, we find that the parameters $r_0$, $\theta_0$ for the hypotenuse $P_1P_3$ are $r_0 = 16k/(5\sqrt{481})$, $\theta_0 = -75/16$ (radians), and for the leg $P_2P_3$ they are $r_0 = 2k/5$, $\theta_0 = 0$. Again, as shown in Appendix 1, the sides of this

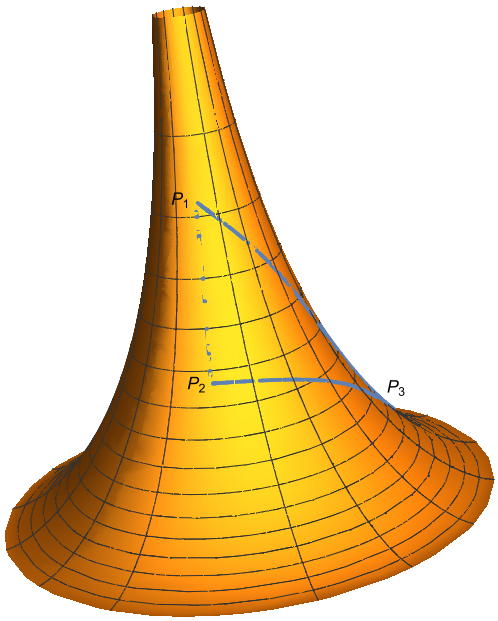


FIGURE 5. The triangle $P_1P_2P_3$ on the pseudo-hemisphere

triangle have the following lengths:

$$\begin{aligned}
\overline{P_1P_2} &= k\,\ln\Big(\frac{\frac{2k}{5}}{\frac{k}{5}}\Big) = k\,\ln(2) \approx 0.693147k\,,\\
\overline{P_1P_3} &= \frac{k}{2}\,\ln\Big(\frac{(k+r_0(\theta_2-\theta_0))(k-r_0(\theta_1-\theta_0))}{(k-r_0(\theta_2-\theta_0))(k+r_0(\theta_1-\theta_0))}\Big)\\
&= \frac{k}{2}\ln\Big(\frac{553+25\sqrt{481}}{72}\Big) \approx 1.36379k\,,\\
\overline{P_2P_3} &= \frac{k}{2}\ln\Big(\frac{\big(k+(2k/5)(2+75/16)\big)\big(k-(2k/5)(75/16)\big)}{\big(k-(2k/5)(2+75/16)\big)\big(k+(2k/5)(75/16)\big)}\Big)\\
&= k\,\ln(3) \approx 1.09861k\,.
\end{aligned}$$

That this is a right triangle follows from the Pythagorean relation $\cosh(\overline{P_1P_3}/k) = \cosh(\overline{P_1P_2}/k)\,\cosh(\overline{P_2P_3}/k)$, which is the identity

$$\cosh\left(\frac{1}{2}\ln\Big(\frac{553+25\sqrt{481}}{72}\Big)\right) = \cosh\big(\ln(2)\big)\,\cosh\big(\ln(3)\big) = \frac{25}{12}\,.$$

The figure in the parameter space $(r, \theta)$ that generates this triangle is shown on the left in Fig. 4, and the projection of that triangle onto the equatorial disk of the pseudo-hemisphere is shown on the right. The triangle itself is shown in Fig. 5.

Let us transport the vector

$$\frac{\partial \boldsymbol{r}}{\partial \theta} = (-r\cos\theta, r\sin\theta, 0)$$

around this triangle from $P_1$ to $P_2$ to $P_3$, and back to $P_1$.

In order to illustrate the general procedure to be followed, we shall take an unnecessarily complicated parametrization of the leg from $P_1$ to $P_2$, using arc length $s$ as the parameter. (The simpler procedure of parameterizing by $r$ is left to the reader in Problem 4.7.)

$$\boldsymbol{\gamma}(s) = \Big(\frac{k}{5}e^{s/k}, 0\Big)\,, \quad 0 \le s \le k\ln(2)\,,$$

that is,

$$s = k\ln\Big(\frac{5r}{k}\Big)\,,$$

so that

$$\boldsymbol{\gamma}'(s) = \Big(\frac{1}{5}e^{s/k}, 0\Big) = \Big(\frac{r}{k}, 0\Big)\,.$$

Consequently, by the formulas given above for the Christoffel symbols,

$$\begin{aligned}
\nabla_{\boldsymbol{\gamma}'(s)}\Big(a(s)\frac{\partial}{\partial r} + b(s)\frac{\partial}{\partial \theta}\Big) &= \frac{1}{5}e^{s/k}\frac{\partial}{\partial r}\Big(a(s)\frac{\partial}{\partial r} + b(s)\frac{\partial}{\partial \theta}\Big) \\
&= \frac{1}{5}e^{s/k}\frac{ds}{dr}\Big(a'(s)\frac{\partial}{\partial r} + b'(s)\frac{\partial}{\partial \theta}\Big) + \frac{1}{5}e^{s/k}\Big(a(s)\frac{\partial^2}{\partial r^2} + b(s)\frac{\partial^2}{\partial r\,\partial \theta}\Big) \\
= \frac{k}{5r}e^{s/k}\Big(a'(s)\frac{\partial}{\partial r} + b'(s)\frac{\partial}{\partial \theta}\Big) &+ \frac{1}{5}e^{s/k}\Big(\Big(a(s)\Gamma^1_{11} + b(s)\Gamma^1_{12}\Big)\frac{\partial}{\partial r} + \Big(a(s)\Gamma^2_{11} + b(s)\Gamma^2_{12}\Big)\frac{\partial}{\partial \theta}\Big) \\
&= \Big(\frac{k}{5r}e^{s/k}a'(s) - \frac{1}{5}e^{s/k}\frac{a(s)}{r}\Big)\frac{\partial}{\partial r} + \Big(\frac{k}{5r}e^{s/k}b'(s) + \frac{1}{5}e^{s/k}\frac{b(s)}{r}\Big)\frac{\partial}{\partial \theta} \\
&= \frac{1}{5r}e^{s/k}\Big(\big(ka'(s) - a(s)\big)\frac{\partial}{\partial r} + \big(kb'(s) + b(s)\big)\frac{\partial}{\partial \theta}\Big)\,.
\end{aligned}$$

Setting both components equal to zero and solving the resulting differential equations, we see that, since $a_0 = 0$ and $b_0 = k/5$

$$a(s) = a_0 e^{s/k} \equiv 0\,, \quad b(s) = b_0 e^{-s/k} = \frac{k}{5}e^{-s/k}\,.$$

We now have the parallel transport of the vector $(0, 1)$ to any point of $P_1P_2$. The squared length of the vector $(a, b)$ is

$$\Big|a\frac{\partial \boldsymbol{r}}{\partial r} + b\frac{\partial \boldsymbol{r}}{\partial \theta}\Big|^2 = E\,a^2 + 2F\,ab + G\,b^2\,.$$

In the present case $E = k^2/r^2$, $F = 0$, and $G = r^2$, so that the length of the vector $(a, b)$ is $\sqrt{(ak/r)^2 + (br)^2}$. Along $P_1P_2$, we have $a = 0$, $b = \frac{k}{5}e^{-s/k}$, $r = \frac{k}{5}e^{s/k}$, so that the length of this vector is $br = k/5$ at each point. That is, the vector does not change its length under this parallel transport.

The side $P_2P_3$ has the angle $\theta$ as parameter:

$$\boldsymbol{\gamma}(\theta) = \Big(\frac{2k}{\sqrt{25 - 4\theta^2}}, \theta\Big)\,, \quad 0 \le \theta \le 2\,,$$

so that

$$\boldsymbol{\gamma}'(\theta) = \Big(\frac{8k\theta}{(25-4\theta^2)^{3/2}}, 1\Big)$$

Using the Christoffel symbols given above, we compute that

$$\nabla_{\boldsymbol{\gamma}'(\theta)}\big(a(\theta), b(\theta)\big) = \Big(a'(\theta) - \frac{4\theta}{25-4\theta^2}a(\theta) - \frac{8k}{(25-4\theta^2)^{3/2}}b(\theta)\Big)\frac{\partial}{\partial r} + \Big(b'(\theta) + \frac{\sqrt{25-4\theta^2}}{2k}a(\theta) + \frac{4\theta}{25-4\theta^2}b(\theta)\Big)\frac{\partial}{\partial \theta}\,.$$

We thus have the following system of two initial-value problems:

$$\begin{aligned} a'(\theta) - \frac{4\theta}{25-4\theta^2}a(\theta) &= \frac{8k}{(25-4\theta^2)^{3/2}}b(\theta)\,, \quad a(0) = 0\,, \\ b'(\theta) + \frac{4\theta}{25-4\theta^2}b(\theta) &= -\frac{\sqrt{25-4\theta^2}}{2k}a(\theta)\,, \quad b(0) = \frac{1}{2}\,. \end{aligned}$$

We remark that the second of these equations implies that $b'(0) = 0$.

Using the standard method of solving the first of these linear equations (see Appendix 5), we find that

$$a(\theta) = \frac{1}{\sqrt{25-4\theta^2}}\int_0^\theta \frac{8kb(t)}{25-4t^2}\,dt\,.$$

Substituting this value into the second equation and differentiating with respect to $\theta$ yields the following second-order initial-value problem for $b(\theta)$:

$$b''(\theta) + \frac{4\theta}{25-4\theta^2}b'(\theta) + \frac{200}{(25-4\theta^2)^2}b(\theta) = 0\,, \quad b(0) = \frac{1}{2}\,, \quad b'(0) = 0\,.$$

A technique for solving this last equation can be found in Appendix 5. (See Problem 5.12 of that appendix. Logically, all we need to do is verify a solution when it is presented to us. This we now do: The solution is

$$b(\theta) = \frac{25-4\theta^2}{50}\,,$$

so that

$$a(\theta) = \frac{4k\theta}{25\sqrt{25-4\theta^2}}\,.$$

Again, we note that the length of the vector $\big(a(\theta), b(\theta)\big)$, which is the square root of $\big(a(\theta)k/r\big)^2 + \big(rb(\theta)\big)^2$, has the constant value $k/5$ along the curve. Upon arriving at $P_3$, our vector, which began as $(0,1)$ at $P_1$ and became $(0,1/2)$ at $P_2$, now has the value

$$\big(a(2), b(2)\big) = \Big(\frac{8k}{75}, \frac{9}{50}\Big)\,.$$

For the last leg of its journey, we send this vector back to $P_1$ along $P_3P_1$. The equation of this curve is

$$r = \frac{16k}{\sqrt{12025-(16\theta+75)^2}} = \frac{4k}{\sqrt{400-150\theta-16\theta^2}}\,.$$

Thus, with the parametrization $\theta \mapsto \boldsymbol{\gamma}(\theta) = \big(r(\theta), \theta\big)$, we find that

$$\boldsymbol{\gamma}'(\theta) = \big(A(\theta), 1\big)\,,$$

where

$$A(\theta) = \frac{4k(16\theta + 75)}{(400 - 150\theta - 16\theta^2)^{3/2}}\,.$$

As before, we get a system of two linear first-order differential equations for the vector field $\big(a(\theta), b(\theta)\big)$ that results from translating our initial vector along $P_3P_1$:

$$\begin{aligned} a'(\theta) &= \frac{A(\theta)a(\theta)}{r(\theta)} + \frac{\big(r(\theta)\big)^3 b(\theta)}{k^2} \\ &= \frac{(16\theta + 75)a(\theta)}{400 - 150\theta - 16\theta^2} + \frac{64kb(\theta)}{(400 - 150\theta - 16\theta^2)^{3/2}}\,, \\ b'(\theta) &= -\frac{A(\theta)b(\theta)}{r(\theta)} - \frac{a(\theta)}{r(\theta)} \\ &= -\frac{(16\theta + 75)b(\theta)}{400 - 150\theta - 16\theta^2} - \frac{\sqrt{400 - 150\theta - 16\theta^2}\,a(\theta)}{4k}\,. \end{aligned}$$

These equations and the conditions $a(2) = 8k/75$, $b(2) = 9/50$ imply that $b'(2) = -139/200$.

As in the previous transport, we easily find an integrating factor for the equation satisfied by $a(\theta)$, namely $\sqrt{400 - 150\theta - 16\theta^2}$, and we solve it explicitly, with the initial condition $a(2) = 8k/75$, getting

$$a(\theta) = \frac{16k}{25\sqrt{400 - 150\theta - 16\theta^2}} + \frac{1}{\sqrt{400 - 150\theta - 16\theta^2}} \int_2^\theta \frac{64kb(t)}{400 - 150t - 16t^2}\,dt\,.$$

Again, as before, substituting this value of $a(\theta)$ into the equation for $b(\theta)$ and differentiating leads to the second-order linear equation

$$b''(\theta) + \frac{(16\theta + 75)b'(\theta)}{400 - 150\theta - 16\theta^2} + \frac{24050b(t)}{(400 - 150\theta - 16\theta^2)^2} = 0\,.$$

We can now call on *Mathematica* to solve these equations for us, and it does not fail. It tells us that

$$\begin{aligned} a(\theta) &= \frac{k}{48100}\left(\frac{24000 + 5120\theta}{\sqrt{400 - 150\theta - 16\theta^2}} - 576\right), \\ b(\theta) &= \frac{1}{48100}\big(80(400 - 150\theta - 16\theta^2) + (675 + 144\theta)\sqrt{400 - 150\theta^2 - 16\theta^2}\big)\,. \end{aligned}$$

These formulas are certainly more complicated than those we dealt with in translating the vector over the other two legs. But it is not difficult to verify that the length of the vector continues to be $k/5$, as *Mathematica* computes the value of

$$\big(ka(\theta)/r(\theta)\big)^2 + \big(b(\theta)r(\theta)\big)^2 = \frac{k^2}{25}\,.$$

When the vector arrives back at $P_1$ after its journey, it is found to be the vector $\big(a(0), b(0)\big) = \big(12k/925, 35/37\big)$. We note that the basis vectors for the tangent space at a general point are

$$\frac{\partial \boldsymbol{r}}{\partial r} = \left(\cos\theta, \sin\theta, -\frac{\sqrt{k^2 - r^2}}{r}\right) \text{ and } \frac{\partial \boldsymbol{r}}{\partial \theta} = (-r\sin\theta, r\cos\theta, 0)\,.$$

Thus, the vector that began at the point $r = k/5$, $\theta = 0$ as $(0, 1)$ was $\boldsymbol{u} = (0, k/5, 0)$. The one that arrived at the end of the journey at the same point was

$\boldsymbol{v} = (12k/925, 7k/37, -24\sqrt{6}k/925)$. Both of these vectors have length $k/5$, and the angle $\varphi$ between them satisfies

$$\cos\varphi = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{|\boldsymbol{u}|\,|\boldsymbol{v}|} = \frac{25}{k^2}\frac{7k^2}{185} = \frac{35}{37}\,.$$

We recall (see Chapter 1 and Appendix 1) that the acute angles of triangle $P_1P_2P_3$ are given by the formulas

$$\sin\big(\angle P_1\big) = \frac{\sinh\big(\overline{P_2P_3}/k\big)}{\sinh\big(\overline{P_1P_3}/k\big)}\,,\quad \sin\big(\angle P_3\big) = \frac{\sinh\big(\overline{P_1P_2}/k\big)}{\sinh\big(\overline{P_1P_3}/k\big)}\,.$$

Given the lengths of the sides, it follows that

$$\sinh\big(\overline{P_1P_2}/k\big) = \frac{3}{4}\,,\quad \sinh\big(\overline{P_2P_3}/k\big) = \frac{4}{3}\,,$$
$$\sinh\big(\overline{P_1P_3}/k\big) = \frac{481+25\sqrt{481}}{12\sqrt{1106+50\sqrt{481}}} = \alpha\,.$$

Consequently

$$\begin{aligned}\sin\big(\angle P_1\big) &= \frac{3}{4\alpha}\,,\quad \cos\big(\angle P_1\big) = \frac{\sqrt{16\alpha^2-9}}{4\alpha}\,,\\ \sin\big(\angle P_3\big) &= \frac{4}{3\alpha}\,,\quad \cos\big(\angle P_3\big) = \frac{\sqrt{9\alpha^2-16}}{3\alpha}\,,\end{aligned}$$

and therefore (in a messy computation that *Mathematica* can perform instantly)

$$\sin\big(\angle P_1 + \angle P_3\big) = \frac{35}{37}\,.$$

Comparing this result with the angle $\varphi$ between the initial and final positions of the tangent vector that was transported, we see that

$$\varphi = \frac{\pi}{2} - \angle P_1 - \angle P_3\,,$$

and since $\angle P_2$ is a right angle, finally,

$$\varphi = \pi - (\angle P_1 + \angle P_2 + \angle P_3)\,.$$

To summarize, the angle $\varphi$ between the initial and final positions of the parallel-transported vector is precisely the angle defect of the triangle around which the vector was transported. Obviously, this result is hardly a mere coincidence that occurs only for the triangle we chose. It holds in general. Since, as shown in Chapter 1 and Appendix 1, the angle defect is proportional to the area of the triangle, parallel transport provides a sort of planimeter, measuring the area of any triangle around whose edges the vector is transported.

## 5. Geodesics

A path having a parametrization $\boldsymbol{\gamma}(t) = \big(x^1(t), \ldots, x^n(t)\big)$, $0 \le t \le a$, and such that $\boldsymbol{\gamma}'(t)$ is the parallel transport of $\boldsymbol{\gamma}'(0)$ along $\Pi$ from $\boldsymbol{\gamma}(0)$ to $\boldsymbol{\gamma}(t)$, is called a *geodesic*. Since $\boldsymbol{w}(t)$, the parallel transport of $\boldsymbol{w}(0)$, is defined as the solution of an initial-value problem for a system of ordinary differential equations, and such an equation has a unique solution, it follows that the parametrization of a geodesic is unique. The *image* of the path can be traced with a different parametrization, by the simple device of changing variables, but the reparametrization no longer counts as a geodesic. Intuitively, when the parameter $t$ is thought of as time, the length of

the tangent vector represents the speed. Since that length does not change during parallel transport, the only traversal that counts as a geodesic is one at constant speed. Thus, a geodesic should be thought of as a *process* rather than as an *object*. Informally, we often refer to any "shortest path"[5] as a geodesic; technically, a geodesic is a *traversal of a shortest path at constant linear speed.* As we shall now show, if $s$ is arc length along the path, then the parameter $t$ is necessarily $s/|\boldsymbol{\gamma}'(0)|$, and hence the length of the path from $\boldsymbol{\gamma}(0)$ to $\boldsymbol{\gamma}(t)$ is $s = t|\boldsymbol{\gamma}'(0)|$.

By definition, a geodesic is a path $\boldsymbol{\gamma}(t)$ such that

$$\nabla_{\boldsymbol{\gamma}'(t)}\boldsymbol{\gamma}(t)\big(\boldsymbol{\gamma}'(t)\big) \equiv \mathbf{0}\,,$$

and for this path, Eq. (2) becomes

$$(x^i)'' + \Gamma^i_{jk}(x^j)'(x^k)' \equiv 0\,.$$

But this is precisely the differential equation that we found, using the Euler equations of the calculus of variations, to be satisfied by a path that minimizes arc length. (See Problems 4.14 and 4.15 of Volume 1.)

We can now justify the informal conflation of geodesics with minimal paths that we allowed ourselves in Chapter 4 of Volume 1. Minimal paths and geodesics are both constructed by solving the same differential equation derived using the calculus of variations. Geodesics are crucial in constructing the exponential map (Chapter 6), which implements the intuitive idea that the tangent space is a "straightened out" copy of the manifold, locally. The connection made here with parallel transport shows that they are also the "straightest" paths on the manifold, those for which the tangent line "wiggles" to the minimum possible extent.

## 6. The Whitney Embedding Theorem

So far in this appendix, we have constructed a theory of manifolds in which an $n$-dimensional manifold is a self-contained universe, a geometric object whose nature is unspecified, except that pieces of it are smoothly homeomorphic by means of a collection of charts to open sets in some Euclidean space $\mathbb{R}^n$. Thus far, we are simply following the abstract language of Riemann. But to define any specific manifold, one still needs an object to stand in as the domain of these charts. Where are we to get it, if not from some Euclidean space? It thus appears that Euclidean spaces are not, after all, something we can simply forget about. There is something universal about them. Can anyone visualize "space" without thinking of some real vector space—or, rather, some observable physical objects? It seems doubtful. That this universality of Euclidean space is not just a limitation of our imagination, but is an essential property of any manifold is the content of theorems called *embedding theorems*, introduced in 1935 by Hassler Whitney (1907–1989). These theorems assert that any smooth manifold can be embedded as a submanifold of some Euclidean space. For our purposes we want to establish only the basic principle that there is really no loss in generality in assuming that every manifold is a hypersurface in some high-dimensional Euclidean space. We do not need any lower bound on the dimension of the ambient Euclidean space. To strip the theorem down to the minimum, we shall consider only compact manifolds. Thus we come to the following theorem.

[5] That informal usage also ignores another distinction, namely the distinction between a local minimum, a local maximum, and a local "saddle point." These paths are really points of *stationarity*, where the variation vanishes.

**Theorem 4.5.** *Any compact $C^\infty$ manifold $\mathfrak{M}$ of dimension $n$ can be embedded as a closed submanifold of a Euclidean space $\mathbb{R}^N$ for sufficiently large $N$.*

Before embarking on the proof, we note that, as discussed in connection with sectional curvature in Chapter 6 of Volume 1, embedding as a submanifold requires that the inverse mapping be $C^\infty$, for which it is necessary and sufficient that the Jacobian matrix of the embedding be of full rank (in this case, $n$) at each point. Since the domain of our hypothetical embedding $\varphi : \mathfrak{M} \to \mathbb{R}^N$ is a manifold, however, not a Euclidean space, we need to explain what is meant by its Jacobian. In general, for a mapping $\varphi : \mathfrak{M} \to \mathfrak{N}$ of one manifold into another, we choose a local chart $\theta$ at a point $P \in \mathfrak{M}$ and a second local chart $\psi$ at the point $\varphi(P) \in \mathfrak{N}$. In those coordinates, the Jacobian is the Jacobian of the mapping $\psi \circ \varphi \circ \theta^{-1}$. (The tensor quantity whose coordinate representation is this matrix is merely the differential $d\varphi$ discussed in connection with covariance and contravariance in Chapter 2 of Volume 1.) Since the range in the present case is a Euclidean space $\mathbb{R}^n$, we can take the trivial chart $\psi(\boldsymbol{y}) = \boldsymbol{y} = (y^1, \ldots, y^N)$ on $\mathfrak{N}$. A convenient choice of $\theta$ will appear in the course of the proof. It is an easy exercise to show that the rank of the Jacobian is well defined, that is, depends only on the mapping $\varphi$ and not on the choice of local coordinates at $P$.

The following proof is the easy first part of a more sophisticated proof due to Vipul Naik, a recent (2013) Ph. D. at the University of Chicago:

`www.cmi.ac.in/~vipul/writeupsandpresentations/proofofwhitneyembedding.pdf`

Naik actually shows that $N$ can be made as small as $2n+1$. The full Whitney Embedding theorem asserts that it can be $2n$, but generally not smaller than that.

Proof. We shall show that $N = m(n+1)$ will suffice, where $m$ is the number of open charts needed to cover the compact manifold $\mathfrak{M}$, and we shall without loss of generality assume that the covering is locally finite.

Let $U_1, \ldots, U_m$ be a basic set of domains of charts $\varphi_1, \ldots, \varphi_m$, where $\varphi_j : U_j \to \mathbb{R}^n$ is assumed to have the form $P \mapsto \big(x_j^1(P), \ldots, x_j^n(P)\big)$, and let $f_1, \ldots, f_m$ be a $C^\infty$ partition of unity subordinate to this covering. The mapping $\psi_j$ defined by the relation

$$\psi_j(P) = \begin{cases} f_j(P)\varphi_j(P) & \text{if } P \in U_j\,, \\ 0 & \text{if } P \notin U_j \end{cases}$$

is a $C^\infty$ mapping from all of $\mathfrak{M}$ into $\mathbb{R}^n$.

We then define a mapping $\varphi : \mathfrak{M} \to \mathbb{R}^{mn+m}$ as an $m$-tuple of $(n+1)$-tuples:

$$\varphi(P) = \big(f_1(P), \psi_1(P); f_2(P), \psi_2(P); \ldots; f_m(P), \psi_m(P)\big)\,.$$

It is obvious from its definition that the mapping $\varphi$ is a $C^\infty$ function. What we need to show is that it is one-to-one and that it has rank $n$ at each point.

To show that $\varphi$ is one-to-one, suppose $\varphi(P) = \varphi(Q)$. Then $f_j(P) = f_j(Q)$ for $j = 1, 2, \ldots, m$, and at least one of these is non-zero, since $\{f_j\}$ is a partition of unity. It then follows that $P$ and $Q$ belong to the same set $U_j$, and that $\varphi_j(P) = \varphi_j(Q)$. But $\varphi$, being a coordinate chart, is one-to-one, and therefore $P = Q$.

It remains to be shown that the mapping $\varphi$ has rank $n$ at each point, that is, we need to show that its Jacobian matrix

$$\begin{pmatrix} D_1 f_1 & D_2 f_1 & \cdots & D_n f_1 \\ D_1(x_1^1 f_1) & D_2(x_1^1 f_1) & \cdots & D_n(x_1^1 f_1) \\ D_1(x_1^2 f_1) & D_2(x_1^2 f_1) & \cdots & D_n(x_1^2 f_1) \\ \vdots & \vdots & \ddots & \vdots \\ D_1(x_1^n f_1) & D_2(x_1^n f_1) & \cdots & D_n(x_1^n f_1) \\ D_1 f_2 & D_2 f_2 & \cdots & D_n f_2 \\ D_1(x_2^1 f_2) & D_2(x_2^1 f_2) & \cdots & D_n(x_2^1 f_2) \\ \vdots & \vdots & \ddots & \vdots \\ D_1(x_m^n f_m) & D_2(x_m^n f_m) & \cdots & D_n(x_m^n f_m) \end{pmatrix}.$$

is a one-to-one mapping at each point $P$ where it is evaluated. We have used the $D_i$ notation here for partial derivatives, since we have not yet chosen the chart $\theta$ that would allow us to write $D_i$ as $\partial/\partial x^i$. We are now going to choose that chart.

Let $P$ be a fixed but arbitrary point of $\mathfrak{M}$. Then $f_j(P) > 0$ for at least one index $j$, and without loss of generality, we take $j = 1$. We now specify that the chart $\theta$ used in computing the Jacobian is to be $\varphi_1$. Then the first $n+1$ coordinates of the mapping $\varphi$ are simply given as

$$\left(f_1(x_1^1, \ldots, x_1^n), x_1^1 f_1(x_1^1, \ldots, x_1^n), x_1^2 f_1(x_1^1, \ldots, x_1^n), \ldots x_1^n f_1(x_1^1, \ldots, x_1^n)\right),$$

and the Jacobian matrix is

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1^1} & \frac{\partial f_1}{\partial x_1^2} & \cdots & \frac{\partial f_1}{\partial x_1^n} \\ f_1 + x_1^1 \frac{\partial f_1}{\partial x_1^1} & x_1^1 \frac{\partial f_1}{\partial x_1^2} & \cdots & x_1^1 \frac{\partial f_1}{\partial x_1^n} \\ x_1^2 \frac{\partial f_1}{\partial x_1^1} & f_1 + x_1^2 \frac{\partial f_1}{\partial x_1^2} & \cdots & x_1^2 \frac{\partial f_1}{\partial x_1^n} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^n \frac{\partial f_1}{\partial x_1^1} & x_1^n \frac{\partial f_1}{\partial x_1^2} & \cdots & f_1 + x_1^n \frac{\partial f_1}{\partial x_1^n} \end{pmatrix}.$$

We claim that this $(n+1) \times n$ matrix has rank $n$, so that the rank of the Jacobian is at least $n$. Since it has only $n$ columns, it cannot have rank larger than $n$, and therefore its rank is $n$. To see why the $(n+1) \times n$ matrix just written has rank $n$, note that the rank of a matrix is unaffected when row-reducing operations such as are used in Gaussian elimination is performed. If we subtract $x_1^j$ times the first row from row $j+1$, the result is the matrix

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1^1} & \frac{\partial f_1}{\partial x_1^2} & \cdots & \frac{\partial f_1}{\partial x_1^n} \\ f_1 & 0 & \cdots & 0 \\ 0 & f_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f_1 \end{pmatrix}.$$

Since the last $n$ rows of this matrix constitute the matrix $f_1 I$, where $I$ is the $n \times n$ identity matrix and $f_1$ is non-zero, the matrix has rank $n$. □

## 7. Problems

**Problem 4.1.** Verify that the chart mappings $\psi_{11} \circ \psi_{22}^{-1}$, $\psi_N \circ \psi_{11}^{-1}$ and $\psi_{11} \circ \psi_N^{-1}$ that define $\mathbb{S}^n(r)$ as a manifold are analytic. It follows that $\mathbb{S}^n(r)$ is an analytic ($C^\omega$) manifold.

**Problem 4.2.** Verify that the inverse $\psi_2^{-1}$ for the second chart given on the torus $\mathbb{T}^2$ is the one stated in the text.

**Problem 4.3.** Verify that the sphere $\mathbb{S}^n(r)$ is orientable. To do so, modify the mapping $\psi_S$ to be

$$\widetilde{\psi}_S(\boldsymbol{\xi}) = \Big(\frac{-r\xi^1}{r+\xi^{n+1}}, \frac{r\xi^2}{r+\xi^{n+1}}, \ldots, \frac{r\xi^n}{r+\xi^{n+1}}\Big)\,,$$

and show that the Jacobian of the mapping $\widetilde{\psi}_S \circ \psi_N$ is positive at every point.

**Problem 4.4.** In connection with the topology on the real projective plane $\mathbb{P}^2$, prove that $\tilde{U} \cap \tilde{V} = \tilde{W}$, where $W = (U \cap V) \cup \big(U \cap (-V)\big) \cup \big((-U) \cap V\big) \cup \big((-U) \cap (-V)\big)$. Since $W$ is an open set in $\mathbb{S}^2(1)$ if $U$ and $V$ are, it follows that the intersection of two open sets in $\mathbb{P}^2$ is open. The facts that $\varnothing$, $\mathbb{P}^2$, and the union of any collection of open sets are all open are trivial.

**Problem 4.5.** Consider the charts $\psi_1$, and $\psi_3$ that form part of the differentiable structure on the real projective plane. Since

$$\psi_1 \circ \psi_3^{-1}\big(\psi_3 \circ \psi_1^{-1}(\boldsymbol{y})\big) = \boldsymbol{y}$$

on the domain of this function, which is the open unit disk with the axes removed, that is, the set of $\boldsymbol{y} = (y^1, y^2)$ such that $(y^1)^2 + (y^2)^2 < 1$ and $y^1 \neq 0 \neq y^2$, we must have

$$J(\psi_1 \circ \psi_3^{-1})\big(\psi_3 \circ \psi_1^{-1}(\boldsymbol{y})\big) \cdot J(\psi_3 \circ \psi_1^{-1})(\boldsymbol{y}) = 1\,.$$

Here $J(\psi)$ denotes the Jacobian determinant of any mapping $\psi : \mathbb{R}^n \to \mathbb{R}^n$. Verify this equation.

**Problem 4.6.** Consider the *real projective line* $\mathbb{P}^1$, a point of which is a two-element set whose elements are antipodal points on the unit circle. That is, a point is $\{\boldsymbol{\xi}, -\boldsymbol{\xi}\} = \big\{(\xi^1, \xi^2), (-\xi^1, -\xi^2)\big\}$, where $(\xi^1)^2 + (\xi^2)^2 = 1$. Given the topology analogous to the one on $\mathbb{P}^2$, we define two open sets that cover $\mathbb{P}^1$, namely $U_1 = \big\{\{\boldsymbol{\xi}, -\boldsymbol{\xi}\} : \xi^1 \neq 0\big\}$ and $U_2 = \big\{\{\boldsymbol{\xi}, -\boldsymbol{\xi}\} : \xi^2 \neq 0\big\}$ and charts $\psi_1 : U_1 \to (-1, 1)$ and $\psi_2 : U_2 \to (-1, 1)$ by

$$\begin{aligned} \psi_1\big(\{\boldsymbol{\xi}, -\boldsymbol{\xi}\}\big) &= \operatorname{sgn}(\xi^1)\xi^2 \\ \psi_2\big(\{\boldsymbol{\xi}, -\boldsymbol{\xi}\}\big) &= \operatorname{sgn}(\xi^2)\xi^1\,. \end{aligned}$$

Describe $\psi_1^{-1}(y)$, $\psi_2^{-1}(y)$, $\psi_2 \circ \psi_1^{-1}(y)$, and $\psi_1 \circ \psi_2^{-1}(y)$ explicitly. How smooth is $\mathbb{P}^1$? Is it orientable?

Also, show that the mapping $f\big(\{\boldsymbol{\xi}, -\boldsymbol{\xi}\}\big) = \big((\xi^1)^2 - (\xi^2)^2, 2\xi^1\xi^2\big) = \boldsymbol{\eta}$ is a one-to-one mapping of $\mathbb{P}^1$ onto the unit circle and has inverse

$$\begin{aligned} f^{-1}(\boldsymbol{\eta}) &= f^{-1}(\eta^1, \eta^2) = \\ &= \left\{ \Big(\sqrt{\frac{1+\eta^1}{2}}, \operatorname{sgn}(\eta^2)\sqrt{\frac{1-\eta^1}{2}}\Big), \Big(-\sqrt{\frac{1+\eta^1}{2}}, -\operatorname{sgn}(\eta^2)\sqrt{\frac{1-\eta^1}{2}}\Big) \right\}. \end{aligned}$$

How smooth are the mappings $f$ and $f^{-1}$?

**Problem 4.7.** In Example 4.9, transport the vector $(0, 1)$ along the leg $P_1P_2$ of the geodesic triangle using $r$ as the parameter instead of arc length $s$, as was done in the text. (This is much simpler than using $s$.)

**Problem 4.8.** Verify the equation

$$X(f) = \boldsymbol{X} \cdot \nabla \tilde{f}\,.$$

**Problem 4.9.** A vector $\boldsymbol{a} = (a^1, a^2, a^3)$ in $\mathbb{R}^3$ can be naturally associated with a skew-symmetric $3 \times 3$ matrix

$$\boldsymbol{A} = \begin{pmatrix} 0 & -a^3 & a^2 \\ a^3 & 0 & -a^1 \\ -a^2 & a^1 & 0 \end{pmatrix}.$$

Show that, if $\boldsymbol{b} = (b^1, b^2, b^3)$ is associated in this way with the matrix $\boldsymbol{B}$, then the cross product $\boldsymbol{a} \times \boldsymbol{b}$ is associated with $[\boldsymbol{A}, \boldsymbol{B}] = \boldsymbol{A}\,\boldsymbol{B} - \boldsymbol{B}\,\boldsymbol{A}$. (Replacing an associative product with its commutator, that is, replacing $AB$ with $[A, B]$ is a standard way of turning an associative algebra into a Lie algebra. If the algebra happens to be commutative, of course, the Lie algebra is trivial, since all products are equal to zero.)

**Problem 4.10.** Verify that the triangle $P_1P_2P_3$ on the pseudo-hemisphere is a right triangle by showing that

$$\cosh\left(\frac{1}{2}\ln\left(\frac{553 + 25\sqrt{481}}{72}\right)\right) = \cosh\big(\ln(2)\big)\,\cosh\big(\ln(3)\big) = \frac{25}{12}\,.$$

## APPENDIX 5

# Differential Equations

A vector field on a manifold is the solution of a system of differential equations. Since it is not obvious that a differential equation has a solution, we need an existence theorem to assure us that we are not working in a vacuum when we use manifolds, that is, that we can create a vector field from a given point having desired properties. We also need a uniqueness theorem to assure us that our creative work produces an unambiguous result. The theorem stated below in terms of functions whose values are in $\mathbb{R}^p$, will suffice for our purposes.

*Notation:* The *norm* or *length* of a vector $\boldsymbol{x} = (x^1, x^2, \ldots, x^n)$ in $\mathbb{R}^n$ is the non-negative real number

$$|\boldsymbol{x}| = \sqrt{(x^1)^2 + \cdots + (x^n)^2}\,.$$

We shall adopt a regrettably confusing notation for differentials, defining

$$d\boldsymbol{x} = dx^1 \cdots dx^n\,.$$

The only use we shall have for this notation in this section is in integration over a region of $\mathbb{R}^n$, and it is a mere convenience to make our formulas look better. The reader must remember that, in this context, the notation $d\boldsymbol{x}$ does *not* mean what it might very well mean when $\boldsymbol{x}$ is a function of variables $u^1, \ldots, u^k$, namely

$$\frac{\partial \boldsymbol{x}}{\partial u^1}\, du^1 + \cdots + \frac{\partial \boldsymbol{x}}{\partial u^k}\, du^k\,.$$

If $\boldsymbol{y}(\boldsymbol{x}) = \big(y^1(x^1, \ldots, x^n), \ldots, y^p(x^1, \ldots, x^n)\big)$ is a vector-valued function with values in $\mathbb{R}^p$, its partial derivatives and integrals are defined componentwise, as vector-valued functions of $\boldsymbol{x}$ in the case of the derivatives and as a vector-valued function of the region of integration in the case of integrals

$$\begin{aligned} \frac{\partial \boldsymbol{y}}{\partial x^j} &= \Big(\frac{\partial y^1}{\partial x^j}, \ldots, \frac{\partial y^p}{\partial x^j}\Big)\,, \\ \int\limits_R \boldsymbol{y}(\boldsymbol{x})\, d\boldsymbol{x} &= \Big(\int\limits_R y^1(\boldsymbol{x})\, d\boldsymbol{x}\,,\ \ldots\,,\ \int\limits_R y^p(\boldsymbol{x}))\, d\boldsymbol{x}\Big)\,. \end{aligned}$$

The fundamental inequality

$$\Big|\int\limits_R \boldsymbol{y}(\boldsymbol{x})\, d\boldsymbol{x}\Big| \le \int\limits_R |\boldsymbol{y}(\boldsymbol{x})|\, d\boldsymbol{x}\,,$$

where the right-hand side is an ordinary multiple integral of a real-valued function, and the fundamental theorem of calculus

$$\frac{d}{dt}\int_a^t \boldsymbol{y}(p)\, dp = \boldsymbol{y}(t)$$

for vector-valued functions are very easy to prove and are left as an exercise (Problem 5.1 below).

## 1. The Initial-value Problem

The initial-value problem of first order is determined by a continuous function $\boldsymbol{f}(x, \boldsymbol{y})$ mapping a connected open set $U \subseteq \mathbb{R}^{p+1}$ into $\mathbb{R}^p$ together with a point $(x_0, \boldsymbol{y}_0) = (x_0, y_0^1, \ldots, y_0^p)$ of $U$. Solving the initial-value problem means determining a continuously differentiable vector-valued function $\boldsymbol{y}(x)$ for $x$ in an open interval containing $x_0$ and satisfying the following conditions

$$\boldsymbol{y}'(x) = \boldsymbol{f}\big(x, \boldsymbol{y}(x)\big) \tag{3}$$

$$\boldsymbol{y}(x_0) = \boldsymbol{y}_0\,. \tag{4}$$

When we assume no more about $\boldsymbol{f}$ than its continuity, this problem can be difficult to solve; indeed, it is not obvious that a solution even exists. Fortunately, the functions dealt with in physics are much better than general continuous functions.

**Theorem 5.1.** *Assume that for each $j = 1, \ldots, p$ the function $\boldsymbol{f}$ has a partial derivative with respect to $y^j$ on a $(p+1)$-dimensional cube $C$ of side $2\delta > 0$ with center at $(x_0, \boldsymbol{y}_0)$ and that these partial derivatives are all bounded on $C$, that is, for some positive number $M$ and all $\boldsymbol{x} = (x, y^1, \ldots, y^p)$ satisfying $x_0 - \delta \le x \le x_0 + \delta$, $(y_0^i - \delta \le y^i \le y_0^i + \delta$, $i = 1, 2, \ldots, p$, we have*

$$\Big|\frac{\partial \boldsymbol{f}}{\partial y^j}\Big| \le M\,.$$

*Then the initial-value problem (3)-4) has a unique solution $\boldsymbol{y}(x)$ defined for $x$ sufficiently near $x_0$.*

Proof. Without any loss of generality, we may assume that $M \ge 1$ and $0 < \delta \le 1$, and we may further assume that $M$ is so large that $|\boldsymbol{f}(x, \boldsymbol{y})| \le M$ for all $(x, \boldsymbol{y}) \in C$.

The idea of the proof is to rewrite the differential equation as

$$\boldsymbol{y}(x) = \boldsymbol{y}_0 + \int_{x_0}^{x} \boldsymbol{f}\big(t, \boldsymbol{y}(t)\big)\, dt\,. \tag{5}$$

A function $\boldsymbol{y}(x)$ satisfying this equation will obviously satisfy the initial condition, and differentiation will show that it also satisfies the differential equation. Its continuous differentiability is guaranteed by the expression on the right-hand side of the equation. Conversely, any function having a continuous derivative and satisfying the differential equation and the initial condition will also satisfy Eq. (5).

The existence of such a function follows from a recursive procedure known as the *method of successive approximations*. We start with a function $\boldsymbol{y}_0(x)$, defined to have the constant value $\boldsymbol{y}_0$ for all $x$. We then define $\boldsymbol{y}_{n+1}(x)$ for all $n \ge 0$ by the equation

$$\boldsymbol{y}_{n+1}(x) = \boldsymbol{y}_0 + \int_{x_0}^{x} \boldsymbol{f}\big(t, \boldsymbol{y}_n(t)\big)\, dt\,.$$

All that remains is to show that the sequence $\{\boldsymbol{y}_n(x)\}_{n=0}^{\infty}$ is properly defined in this way—that is, at least for $x$ sufficiently near $x_0$, each point $\big(x, \boldsymbol{f}(y(x)\big)$ belongs to $C$—and that this sequence converges to a continuous function $\boldsymbol{y}(x)$ satisfying Eq. (3). In fact, it will turn out that the condition "$x$ sufficiently near $x_0$" can be the inequality $|x - x_0| \le \delta/(2pM) = \varepsilon$. To prove convergence we shall rely on a

few facts about uniform convergence that are in standard textbooks of advanced calculus (elementary real analysis).

We begin by proving the simple inequality

$$|\boldsymbol{f}(x, \boldsymbol{y}_1) - \boldsymbol{f}(x, \boldsymbol{y}_2)| \leq pM|\boldsymbol{y}_1 - \boldsymbol{y}_2|$$

for all pairs $(x, \boldsymbol{y}_1) \in C$, $(x, \boldsymbol{y}_2) \in C$.

It is really this inequality that we need, not the stronger assumption from which we shall derive it, that the partial derivatives are bounded. An inequality of this type is known as a *Lipschitz condition* after Rudolph Lipschitz (1832–1903). Let $\boldsymbol{y}_1 = (y_1^1, y_1^2, \ldots, y_1^p)$ and $\boldsymbol{y}_2 = (y_2^1, y_2^2, \ldots, y_2^p)$. Then

$$\begin{aligned}
\boldsymbol{f}(x, \boldsymbol{y}_1) - \boldsymbol{f}(x, \boldsymbol{y}_2) &= \boldsymbol{f}(x, y_1^1, y_1^2, \ldots, y_1^{p-1}, y_1^p) - \boldsymbol{f}(x, y_2^1, y_2^2, \ldots, y_2^{p-1}, y_2^p) \\
&= \big(\boldsymbol{f}(x, y_1^1, y_1^2, \ldots, y_1^{p-1}, y_1^p) - \boldsymbol{f}(x, y_2^1, y_1^2, \ldots, y_1^{p-1}, y_1^p)\big) + \\
&+ \big(\boldsymbol{f}(x, y_2^1, y_1^2, \ldots, y_1^{p-1}, y_1^p) - \boldsymbol{f}(x, y_2^1, y_2^2, \ldots, y_1^{p-1}, y_1^p)\big) + \\
&\quad + \cdots + \big(\boldsymbol{f}(x, y_2^1, y_2^2, y_2^3, \ldots, y_2^{p-1}, y_1^p) - \boldsymbol{f}(x, y_2^1, y_2^2, y_2^3, \ldots, y_2^{p-1}, y_2^p)\big)\,.
\end{aligned}$$

We have here a "telescoping" sum. A typical term here is the second, and we note that by the fundamental theorem of calculus we can write it as

$$\int_{y_1^2}^{y_2^2} \frac{\partial}{\partial s} \boldsymbol{f}\big(x, y_2^1, s, y_1^3, \ldots, y_1^{p-1}, y_1^p\big)\, ds\,.$$

Now by our assumption

$$\Big|\frac{\partial}{\partial s} \boldsymbol{f}\big(x, y_2^1, s, \ldots, y_1^{p-1}, y_1^p\big)| \leq M\,,$$

so that this term has norm at most

$$M|y_1^2 - y_2^2| \leq M|\boldsymbol{y}_1 - \boldsymbol{y}_2|\,.$$

The same argument applies to each of the terms, and we conclude that, as claimed,

$$|\boldsymbol{f}(x, \boldsymbol{y}_1) - \boldsymbol{f}(x, \boldsymbol{y}_2)| \leq pM|\boldsymbol{y}_1 - \boldsymbol{y}_2|\,.$$

That being established, we now observe that for $|x - x_0| \leq \varepsilon$, we have

$$|\boldsymbol{y}_1(x) - \boldsymbol{y}_0| = \Big|\int_{x_0}^{x} \boldsymbol{f}(s, \boldsymbol{y}_0)\, ds\Big| \leq \int_{x_0}^{x} |\boldsymbol{f}(s, \boldsymbol{y}_0)|\, ds \leq M|x - x_0| \leq \frac{\delta}{2p} \leq \frac{\delta}{2} \leq \frac{1}{2}\,,$$

so that $\big(x, \boldsymbol{y}_1(x)\big) \in C$ for all $x$ in this range.

It follows that $\boldsymbol{y}_2(x)$ is well defined, that is $\boldsymbol{f}\big(x, \boldsymbol{y}_1(x)\big)$ is defined and continuous for $x$ in this range. Repetition of the same argument shows that

$$|\boldsymbol{y}_2(x) - \boldsymbol{y}_0| = \Big|\int_{x_0}^{x} \boldsymbol{f}\big(s, \boldsymbol{y}_1(s)\big)\, ds\Big| \leq \int_{x_0}^{x} |\boldsymbol{f}\big(s, \boldsymbol{y}_1(s)\big)|\, ds \leq M|x - x_0| \leq \delta/2p \leq \delta\,.$$

In this way, we get a complete induction showing that each $\boldsymbol{y}_n(x)$ is well defined for $|x - x_0| \leq \varepsilon$ and is a continuous function of $x$.

As for the convergence, we observe that for each $x$ with $|x - x_0| \le \varepsilon$ we have

$$\begin{aligned} |\boldsymbol{y}_{n+1}(x) - \boldsymbol{y}_n(x)| &\le \Big| \int_{x_0}^{x} \boldsymbol{f}\big(t, \boldsymbol{y}_n(t)\big) - \boldsymbol{f}\big(t, \boldsymbol{y}_{n-1}(t)\big)\, dt \Big| \\ &\le \int_{x_0}^{x} \big| \boldsymbol{f}\big(t, \boldsymbol{y}_n(t)\big) - \boldsymbol{f}\big(t, \boldsymbol{y}_{n-1}(t)\big) \big|\, dt \\ &\le pM|x - x_0| \sup_{|t-x_0|\le|x-x_0|} |\boldsymbol{y}_n(t) - \boldsymbol{y}_{n-1}(t)| \\ &\le \frac{1}{2} \sup_{|t-x_0|\le|x-x_0|} |\boldsymbol{y}_n(t) - t, \boldsymbol{y}_{n-1}(t)| \,. \end{aligned}$$

Since

$$\sup_{|t-x_0|\le|x-x_0|} |\boldsymbol{y}_1(t) - \boldsymbol{y}_0(t)| \le \frac{1}{2} \,,$$

it follows that

$$\sup_{|t-x_0|\le|x-x_0|} |\boldsymbol{y}_{n+1}(t) - \boldsymbol{y}_n(t)| \le \frac{1}{2^{n+1}} \,.$$

It follows from the Weierstrass $M$-test that the series

$$\boldsymbol{y}_0 + \sum_{n=1}^{\infty} \big(\boldsymbol{y}_n(t) - \boldsymbol{y}_{n-1}(t)\big)$$

converges uniformly to a continuous function $\boldsymbol{y}(t)$ on the closed interval $[x_0 - \varepsilon, x_0 + \varepsilon]$.

Because of the inequality satisfied by $\boldsymbol{f}$, this means that $\boldsymbol{f}\big(t, \boldsymbol{y}_n(t)\big)$ also converges uniformly to $\boldsymbol{f}\big(t, \boldsymbol{y}(t)\big)$ on that same interval and that the sequence can be integrated termwise. That establishes Eq. (3), and the proof of the existence of a solution to the initial-value problem is now complete.

To prove uniqueness, suppose there were two functions $\boldsymbol{y}(x)$ and $\boldsymbol{z}(x)$ such that

$$\begin{aligned} \boldsymbol{y}(x) &= \boldsymbol{y}_0 + \int_{x_0}^{x} \boldsymbol{f}\big(t, \boldsymbol{y}(t)\big)\, dt \,, \\ \boldsymbol{z}(x) &= \boldsymbol{y}_0 + \int_{x_0}^{x} \boldsymbol{f}\big(t, \boldsymbol{z}(t)\big)\, dt \,. \end{aligned}$$

We would then have the inequality

$$\sup_{|x-x_0|\le\varepsilon} |\boldsymbol{y}(x) - \boldsymbol{z}(x)| \le Mp\varepsilon \sup_{|t-x_0|\le\varepsilon} |\boldsymbol{y}(t) - \boldsymbol{z}(t)| \le \frac{1}{2} \sup_{|t-x_0|\le\varepsilon} |\boldsymbol{y}(t) - \boldsymbol{z}(t)|.$$

Since the variables $x$ and $t$ appearing under the supremum signs are "dummy" variables, that is, the meaning of the expression is unchanged if they are replaced consistently by any other symbol, this means that

$$\frac{1}{2} \sup_{|s-x_0|\le\varepsilon} |\boldsymbol{y}(s) - \boldsymbol{z}(s)| \le 0 \,.$$

But a norm that is not positive can only be the norm of the zero vector, and that means $\boldsymbol{y}(s) \equiv \boldsymbol{z}(s)$ for $x_0 - \varepsilon \le s \le x_0 + \varepsilon$. □

Some of the details of this proof and consequences of the theorem are explored in the problems at the end of this appendix.

**1.1. Linear equations and dependence on initial conditions.** If the function $\boldsymbol{f}$ is such that $\boldsymbol{f}\big(x, \boldsymbol{y}(x)\big) = M(x)\big(\boldsymbol{y}(x)\big)$, where $M(x)$ is a linear operator for each $x$, it is very easy to show that the solution $\boldsymbol{y}(x)$ is a linear function of the initial condition $\boldsymbol{y}_0$. This is, in fact, a trivial computation, which we omit.

Moreover, in this case, the departure of $\boldsymbol{y}(x)$ from its initial value $\boldsymbol{y}_0 = \boldsymbol{y}(x_0)$ satisfies the bound

$$|\boldsymbol{y}(x) - \boldsymbol{y}_0| \le \frac{\|M\|\,|\boldsymbol{y}_0|}{1 - \|M\|\,|x - x_0|}|x - x_0|\,,$$

where $\|M\|$ is the supremum of the norms $\|M(t)\|$ over the range $|t - x_0| \le |x - x_0|$.

This inequality is proved by imitating the proof of uniqueness just given.

## 2. Anatomy of First-order Differential Equations

The present section contains some pedagogical reflections on the way ordinary differential equations of first order and first degree have traditionally been taught to students in beginning calculus courses. The emphasis in such discussions is understandably on getting computational results, so that the student is urged to notice the patterns in the relatively few such equations that have elementary closed-form solutions. These are generally taught in increasing order of complexity: variables-separable, homogeneous, linear, and exact equations. The end of this progression is usually a discussion of integrating factors. This treatment, in my opinion, starts too far in, ignores the *source* of all such equations in physics, and conceals the essential unity of all five of the topics just mentioned. Here is the way I think the discussion ought to run.

Physicists set great store by their conservation laws: conservation of momentum, conservation of angular momentum, conservation of energy. Each of these asserts that some function $f(x, y, z, \dots)$ of observable quantities $x$, $y$, $z$,..., remains constant during a process. To take a very trivial example, consider the case of a particle in free fall near the surface of the earth, neglecting air resistance and the variability of the acceleration $g$ of gravity. If $y = y(t)$ is the elevation of a particle above the earth's surface at time $t$ and $v = v(t)$ is the velocity of the particle at time $t$, then the total energy $E(y, v)$ of the particle per unit mass is the sum of the potential energy per unit mass $gy$ and the kinetic energy per unit mass $\frac{1}{2}v^2$:

$$E(y, v) = gy + \frac{1}{2}v^2\,.$$

Since energy is conserved, $E\big(y(t), v(t)\big)$ is constant at all times $t$. (That is what *conserved* means!) By differentiating this equation with respect to time, we can get a very simple first-order linear differential equation:

$$0 = \frac{dE}{dt} = g\,\frac{dy}{dt} + v\,\frac{dv}{dt}\,,$$

or, in differential form,

$$0 = dE = g\,dy + v\,dv\,.$$

Normally, however, we go in the other direction. From basic laws of physics such as Newton's $F = ma$, we try to work out a conservation law. That is, we start with the differential equation $g\,dy + v\,dv$ and try to replace it with an equivalent relation between the variables $y$ and $v$, not involving any derivatives. If it is equivalent to an equation $f(y, v) = \text{const.}$, the function $f(y, v)$ is called an *integral* of the equation.

In other words, an integral is a quantity that is conserved whenever its variables are functions of time that satisfy the differential equation.

More generally, given a system of $n-1$ differential equations in $n$ variables, which for reasons that will become clear below, we write as

$$\frac{dx^1}{M_1(x^1,\dots,x^n)} = \cdots = \frac{dx^n}{M_n(x^1,\dots,x^n)}\,,$$

we try to find $n-1$ independent functions $f_1(x^1,\dots,x^n),\dots,\ f_{n-1}(x^1,\dots,x^n)$ such that $f_j\big(x^1(t),\dots,x^n(t)\big)$ is a constant $c_j$, $j=1,2,\dots,n-1$, when the differential equations hold.

Theoretically, if we find them, the derivatives will have been eliminated and the system replaced by a system of $n-1$ equations in $n$ variables, allowing us to express all the variables in terms of just one of them. From a geometric point of view, the conserved quantities $f_j(x^1,\dots,x^n)$ "usually" (generically, as algebraic geometers say) provide us with a system of $n-1$ manifolds, each of dimension $n-1$, and characterized as the graphs of the equations $f_j(x^1,\dots,x^n)=\text{const}_j$. in $\mathbb{R}^n$. If the integrals are truly independent, the intersection of these manifolds is a one-dimensional manifold that can be parameterized by time $t$. Thus, as far as the theoretical subject of differential equations is concerned, an equation has been solved when these $n-1$ conservation laws have been found, and the problem is therefore reduced to finding them. Whether one can actually solve the resulting equations need not concern the specialist in differential equations, even though it may be of great importance to the physicist.

Of course, solving a general system of $n-1$ equations in $n$ unknowns rapidly becomes an unmanageable task as $n$ increases, unless the functions appearing in the equations are extremely simple. One hardly knows where to begin with such a system, and the general procedure by which the conservation laws determine the solution remains mysterious. There is one case, however, in which a systematic approach has been found. That is the case when the divergence of the vector $(M_1,\dots,M_n)$ vanishes, that is,

$$\frac{\partial M_1}{\partial x^1} + \cdots + \frac{\partial M_n}{\partial x^n} \equiv 0\,.$$

In that case, if an integral $f(x^1,\dots,x^n)$ has been found, and the equation $f(x^1,\dots,x^n)=k$ can be solved for, say $x^n$, the expression $dx^n/M_n(x^1,\dots,x^n)$ can be omitted, the remaining equations can be multiplied by $\partial f/\partial x^n$ (with $x^n$ replaced by its expression in terms of the other variables after the differentiation is performed), and the reduced system will satisfy the divergence criterion in the remaining variables. This technique is known as the *Jacobi last-multiplier* method, since it implies that after $n-2$ integrals have been used to reduce the system to one equation in two variables, that remaining equation is exact. The details are given below, but first we shall look at the standard types of differential equations discussed in calculus books.

**2.1. Variables separable.** The equation $dE=0$ for the case of a particle in free fall is so simple that it falls under the first class of equations usually studied, those with "variables separable." That means it can be solved by evaluating two integrals such that the integrand is independent of $v$ in one of them and independent of $y$ in the other. (In the present case, one of these integrands is independent of both $y$

and $v$.)

$$E = \int dE = \int g\,dy + \int v\,dv = gy + \frac{1}{2}v^2 + C\,,$$

where the constant $C$ has to be determined from some initial or boundary condition. Potential energy is determined only up to an additive constant, so we can say that $gy + C$ is the potential energy, and then agree that it is to be zero when $y = 0$, thus getting $C = 0$.

Since this is the first kind of differential equation studied, textbooks seldom bother to point out later on that any equation of the form $a(x)\,dx + b(y)\,dy = 0$ is also an *exact* differential equation. That is, there exists a function $f(x, y)$, namely $\int a(x)\,dx + \int b(y)\,dy$ such that the equation says

$$0 = df(x, y) = \frac{\partial f(x, y)}{\partial x}\,dx + \frac{\partial f(x, y)}{\partial y}\,dy\,.$$

In general, since the partial derivatives depend on both $x$ and $y$, we have to be a little careful to verify that the differential form on the right is the differential of some function. That is done by using the test of equality of mixed partial derivatives. In general, however, we do not get our differential equations by differentiating conservation laws. It is rather the reverse that happens: Some real-world model suggests a differential equation, and we derive a conservation law by solving that equation.

**Example 5.1.** Consider the Lotka–Volterra model[1] for the interaction of a predator population $x$ with a prey population $y$. We assume a very simple model in which the predator population is always at the brink of starvation. That is, since we ignore any higher species on the food chain, the population $x$ is always very close to the carrying capacity of its environment, which in turn depends on the size of the prey population. Under those circumstances, an increase in the population of predators actually decreases the rate of growth of that population (since it leads to starvation). Taken in isolation, this would give us the differential equation $dx/dt = -Bx$, where $B$ is a positive constant. Likewise, assuming the prey population $y$ is kept down by the predators, this population has an essentially limitless food supply, so that, again taken in isolation, we should have the Malthusian exponential growth (geometric increase): $dy/dt = Cy$, On the other hand, it is natural to assume that encounters of members of one species with members of the other are proportional to the product of the two populations $xy$. Such encounters tend to increase the food supply for the predator population, and hence also its reproductive rate, while of course they directly decrease the prey population and hence also its reproductive rate. Thus we get two new parameters $A$ and $D$, such that the following *Lotka–Volterra* equations hold:

$$\begin{aligned} \frac{dx}{dt} &= Axy - Bx\,, \\ \frac{dy}{dt} &= Cy - Dxy\,. \end{aligned}$$

From these two equations, we get a single differential equation

$$(Cy - Dxy)\,dx + (Bx - Axy)\,dy = 0\,.$$

[1] Named for Alfred J. Lotka (1860–1949) and Vito Volterra (1860–1940).

Since in general

$$\frac{\partial}{\partial y}(Cy - Dxy) = C - Dx \neq B - Ay = \frac{\partial}{\partial x}(Bx - Axy)\,,$$

this equation is not an exact differential equation.

If we divide it by $xy$, however, we get the variables-separable equation

$$\Big(\frac{C}{x} - D\Big)\,dx + \Big(\frac{B}{y} - A\Big)\,dy = 0\,,$$

and thus a curious "conservation law" for the two populations:

$$x^C y^B = Ke^{Dx+Ay}\,.$$

Of course, all five of the parameters $A$, $B$, $C$, $D$, and $K$ in this law are theoretical and would have to be determined specifically for each given pair of populations over a period of time. Since the environment contains so many variable factors—the weather being an important one, for example—this is not a practical program, and you will probably not find anyone invoking this "law." The true value of this oversimplified model is theoretical: It predicts a periodic fluctuation of the two populations, representable as the traversal of an ovoid trajectory in the $xy$-plane parameterized by time $t$, with the values of $x(t)$ and $y(t)$ being essentially out of phase with each other, like the sine and cosine functions.

**2.2. Exact equations.** An exact differential equation

$$0 = df(x,y) = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy\,,$$

is not particularly difficult to solve.

We can hold $y$ fixed and integrate the first term, for example, getting

$$f(x,y) = \int \frac{\partial f}{\partial x}\,dx + C(y)\,,$$

where the "constant" of integration depends on $y$.

Then if we differentiate with respect to $y$, we get a simple differential equation for $C(y)$:

$$C'(y) = \frac{\partial f}{\partial y} - \int \frac{\partial^2 f}{\partial y\,\partial x}\,dx\,.$$

Here, the first term on the right-hand side is known, being the coefficient of $dy$ in the original equation; the second term is obtained by integrating the $y$-derivative of the coefficient of $dx$ in the original equation. The right hand side is guaranteed to be independent of $x$ by the equality of mixed partial derivatives. Thus, if we integrate again, we will get the function $f(x,y)$ plus a constant $k$ that really is constant this time.

As an example, consider the equation

$$(x^2 - y^2)\,dx + \big(\cos(y) - 2xy\big)\,dy = 0\,.$$

This equation is exact, since $\partial(x^2 - y^2)/\partial y = -2y = \partial\big(\cos(y) - 2xy\big)/\partial x$. Integrating the first term yields

$$f(x,y) = \frac{1}{3}x^3 - xy^2 + C(y)\,.$$

Then

$$C'(y) = \frac{\partial f}{\partial y} - 2xy = \big(\cos(y) - 2xy\big) - \int (-2y)\,dx = \cos(y)\,,$$

and thus $C(y) = k + \sin(y)$. Altogether then

$$f(x,y) = \frac{1}{3}x^2 - xy^2 + \sin(y) + k\,.$$

Thus, conservation laws emerge from exact equations "by quadrature" (using only integration). What more could we wish? Let us now explore that question.

**2.3. Where do conservation laws come from?** Think of Nature as the repository of many fundamental conservation laws. We attempt to discover Nature's secrets by setting up differential equations and solving them to reveal these fundamental principles of the structure of the universe. The trouble is that the equations we set up, as in the Lotka–Volterra model discussed above, are *not* simply obtained by differentiating these laws, because we don't know the laws in advance. What we get, in short, is not the exact equation

$$0 = df(x,y) = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy\,,$$

but an equivalent relation $\mu(x,y)\,df(x,y) = 0$, that is,

$$\mu(x,y)\frac{\partial f}{\partial x}\,dx + \mu(x,y)\frac{\partial f}{\partial y}\,dy = 0\,.$$

The difficulty is that we get these coefficients as two "lumps" and we can't tell which part of each "lump" is $\mu$. Nature, whether subtly or maliciously, hides its secret structures from us by "smearing" the time derivative of the conserved quantity with the function $\mu(x,y)$, which we must try to discover.

Sometimes, as in the case of the Lotka–Volterra equations, it is easy to see what $\mu(x,y)$ might be—it could be $xy$ in that case[2] —but it is often difficult to do that. The general problem of solving an ordinary differential equation of first order and first degree, that is, an equation of the form

$$M(x,y)\,dx + N(x,y)\,dy = 0\,,$$

is to *find* an *integrating factor* $1/\mu(x,y)$ such that

$$\frac{M(x,y)}{\mu(x,y)}\,dx + \frac{N(x,y)}{\mu(x,y)}\,dy = 0$$

is an exact equation, as revealed by the test

$$\frac{\partial}{\partial y}\left(\frac{M(x,y)}{\mu(x,y)}\right) = \frac{\partial}{\partial x}\left(\frac{N(x,y)}{\mu(x,y)}\right).$$

If we write the equation in the equivalent form as

$$\frac{dx}{P(x,y)} = \frac{dy}{Q(x,y)}\,,$$

where $P(x,y) = \frac{N(x,y)}{\mu(x,y)}$ and $Q(x,y) = -\frac{M(x,y)}{\mu(x,y)}$, the criterion for exactness becomes the simple rule

$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} = 0\,,$$

[2] We say "could be" because an integrating factor is not unique.

that is, the divergence of the vector $\big(P(x,y),Q(x,y)\big)$ is zero.

**2.4. Homogeneous equations.** Besides the variables-separable equations, or, more generally, any exact equation, for which no integrating factor is needed, there are two other types commonly considered in calculus books. The first kind, homogeneous equations, are reducible to equations with variables separable, and again, no integrating factor needs to be found.

If the coefficients $M(x,y)$ and $N(x,y)$ are homogeneous functions of degree $r$, which means that for each real number $t$,

$$M(tx,ty) = t^r M(x,y) \text{ and } N(tx,ty) = t^r N(x,y)\,,$$

it is possible to change variables by setting $t = y/x$, so that $y = tx$ and $dy = t\,dx + x\,dt$.

We then get the equation

$$\begin{aligned} 0 = M(x,tx)\,dx + N(x,tx)\,dy &= x^r M(1,t)\,dx + x^r N(1,t)(t\,dx + x\,dt) = \\ &= x^r\big(M(1,t) + t\,N(1,t)\big)\,dx + x^{r+1}N(1,t)\,dt\,. \end{aligned}$$

Dividing this equation by $x^{r+1}\big(M(1,t) + t\,N(1,t)\big)$ yields an equation with the variables separated:

$$\frac{1}{x}\,dx + \frac{N(1,t)}{M(1,t)+N(1,t)}\,dt = 0\,.$$

It can then be integrated directly and $x$ found explicitly as a function of $t$. We then merely replace $t$ throughout by $y/x$ to get the solution of the original equation.

For example, consider the equation

$$(x^2+y^2)\,dx + xy\,dy = 0\,.$$

Setting $y = tx$, we get

$$0 = x^2(1+t^2)\,dx + x^2 t\,(x\,dt + t\,dx)\,,$$

or,

$$(1+2t^2)\,dx + tx\,dt = 0\,,$$

so that

$$0 = \frac{1}{x}\,dx + \frac{t}{\big(1+2t^2\big)}\,dt\,.$$

It follows that

$$\ln(x) + \frac{1}{4}\ln(1+2t^2) = C\,,$$

or, more elegantly, $x^2(x^2+2y^2) = k$.

**2.5. Linear equations.** So far, the most complicated equation we have considered is the general exact equation. Variables-separable equations are a simpler special case of exact equations, and homogeneous equations (easily recognized by the fact that every term is of the same total degree in the two variables) are reducible to equations with variables separable. Thus, we have not actually needed to look for integrating factors as yet. But there is one very important class of equations, not generally exact, and for which an integrating factor can be found. That class is the linear equations, which is to say, equations of the form

$$\frac{dy}{dx} + P(x)y = Q(x)\,,$$

or, as we write it

$$\big(P(x)y - Q(x)\big)\,dx + dy = 0\,.$$

If $F(x) = \int P(x)\,dx$ is an indefinite integral of $P(x)$, then $F'(x) = P(x)$, and the equation

$$M(x,y)\,dx + N(x,y)\,dy = e^{F(x)}\big(P(x)y - Q(x)\big)\,dx + e^{F(x)}\,dy = 0$$

is exact, since

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}\,.$$

As a matter of fact, the solution can be written explicitly:

$$y(x) = e^{-F(x)} \int Q(x) e^{F(x)}\,dx\,,$$

The indefinite integral here contains an arbitrary constant term. There may also be an arbitrary constant term in the choice of the function $F(x)$, but any constant term in $F(x)$ will cancel out of the equation.

In summary, here is the whole secret to the exposition of differential equations in calculus books: *In general, looking for an integrating factor is a fool's errand. There are almost no equations for which one can find such a factor.* The corollary of that principle is revealing: *Calculus books expound only the most prominent and useful equations for which an integrating factor was found long ago!!* In other words, the standard examples are chosen not because they are more likely to be useful in studying nature, but simply because they are the ones for which we can cope with the computations!

Unfortunately, this metamathematical fact is seldom communicated to the poor student, who consequently feels frustrated when encountering a new equation for which the known techniques do not work.

## 3. Jacobi's last-multiplier principle*

As we noted above, a system of $n$ first-order differential equations in $n+1$ variables

$$\frac{dx^1}{M_1(x^1,\dots,x^{n+1})} = \cdots = \frac{dx^{n+1}}{M_n(x^1,\dots,x^{n+1})}\,, \tag{6}$$

is considered to be solved if one can find $n$ independent integrals, that is, $n$ independent functions whose values are conserved when the variables satisfy the equations.

When $n = 1$, finding the integral is equivalent to finding an integrating factor to make the equation exact. In the general case, if we somehow find an integral, it can be used to reduce the number of variables and equations by one. In practice, the integral often contains algebraic complications that make the reduced system of $n-1$ equations in $n$ variables less tractable than it appeared before. For example, if you have an integral for a system of two equations in three variables, you can use it to get a single equation in two variables. That equation is very likely not an exact equation, however, and one is still faced with the problem of finding an integrating factor for it.

There is one situation that arises in practice in which an integrating factor can be produced by a method known as the *Jacobi last-multiplier principle*. That situation arises when the divergence of the vector $(M_1,\dots,M_{n+1})$ vanishes:

$$\frac{\partial M_1}{\partial x^1} + \cdots + \frac{\partial M_{n+1}}{\partial x^{n+1}} = 0\,.$$

Given this condition, if you have one non-trivial integral that can be used to express $x^{n+1}$ in terms of $x^1, \ldots, x^n$ and thereby remove the last equation from the system, the result will be a new system of $n-1$ equations in $n$ unknowns. Unfortunately, eliminating that last variable generally destroys the beautiful symmetry that caused the divergence to vanish, and the divergence of the new system does *not* vanish. Jacobi noticed, however, that there is a generalized integrating factor that will correct this imbalance and produce a new system with vanishing divergence. In that way—at least, in theory—one need only find $n-1$ integrals. If that many integrals can be found, applying them sequentially to eliminate variables will produce eventually a single equation in two variables whose divergence vanishes. But that is precisely what is meant by an exact equation, and so the solution of the problem will have been reduced to quadratures (evaluating integrals). This is useful in theory, since if the $n$ integrals are all algebraic functions, then the generalized integrating factors are also. As a result, if the coefficients in the original system are algebraic, the solutions will be integrals of algebraic functions. When the solution can be expressed in terms of integrals of algebraic functions, it is said to be *exactly solvable*.

As a method of actually *finding* the solution, the expressions that are produced by the last-multiplier principle may not be particularly enlightening. Still, mathematicians, if not physicists, find it intriguing that some systems of differential equations have solutions that are integrals of algebraic functions, among which are trigonometric, elliptic, hyperelliptic, and general abelian functions, all of which are produced by solving a problem posed by Jacobi in 1832, known as the *Jacobi inversion problem*.[3] (As is well known, trigonometric functions are the inverses of functions defined as the integrals of functions involving the square root of a quadratic polynomial. Elliptic functions are the inverses of integrals of functions involving the square root of a polynomial of degree three or four; hyperelliptic integrals arise when the polynomial is of degree five or six. A general abelian integral[4] has an integrand $R(x, y)$, where $y$ is a function of $x$ defined by a polynomial equation $p(x, y) = 0$.)

To summarize, Jacobi showed that the vanishing of the divergence can be exploited to solve for the variables one at a time, thereby eliminating equations from the system one at a time, and this process can be continued as long as new integrals can be found for the remaining equations, until finally there is only one equation left in two variables, and it is exact. That is the result we now set out to establish. We state it as a formal theorem:

**Theorem 5.2.** *Consider a system of first-order ordinary differential equations*

$$\frac{dx^1}{M_1} = \frac{dx^2}{M_2} = \cdots = \frac{dx^{n+1}}{M_{n+1}}\,, \tag{7}$$

[3] Mathematicians soon realized that the secret of solving this problem was to use multi-variable versions of the theta functions that Jacobi and Abel had introduced to study elliptic functions. Weierstrass and Riemann simultaneously and independently solved the problem in 1856. Applications of multi-variable theta functions to physics eventually began to appear: Carl Gottfried Neumann (1832–1925) used them to describe a coupled harmonic oscillator in his 1858 doctoral dissertation, and Sof'ya Kovalevskaya so impressed the jury of the Paris Academy of Sciences in the 1888 Bordin Prize competition by using these functions to describe an irregular spinning top, that she won the prize.

[4] Named after Niels Henrik Abel (1802–1829).

*where $M_1, \dots, M_{n+1}$ are functions of $x^1, \dots, x^{n+1}$, and assume that the divergence of the system vanishes, that is,*

$$\frac{\partial M_1}{\partial x^1} + \cdots + \frac{\partial M_{n+1}}{\partial x^{n+1}} \equiv 0\,.$$

*Suppose also that the function $f(x^1, \dots, x^{n+1})$ is constant whenever the system of equations is satisfied, say $f(x^1, \dots, x^{n+1}) = k$. Finally, suppose this last equation can be solved for $x^{n+1}$ as a function $x^{n+1} = h(x^1, \dots, x^n)$. Let $M_j^*(x^1, \dots, x^n) = M_j\big(x^1, \dots, x^n, h(x^1, \dots, x^n)\big)$. Then there exists a function $\mu(x^1, \dots, x^n)$ such that the system*

$$\frac{dx^1}{\mu M_1^*} = \cdots = \frac{dx^n}{\mu M_n^*}$$

*also has divergence zero.*

PROOF. Since we will be applying this method in a case where all of the variables $x^j$ depend on time, we shall write the proof with that in mind, although logically we do not need to do so, and we could regard $x^1, \dots, x^{n+1}$ as pure variables. Suppose that we have found an integral $f(x^1, \dots, x^{n+1})$ such that if $x^1(t), \dots, x^{n+1}(t)$ are functions of time satisfying Eq. (7), the function $\varphi(t) = f\big(x^1(t), \dots, x^{n+1}(t)\big)$ has the constant value $k$ for all $t$. Without loss of generality, assume that the equation $f(x^1, \dots, x^{n+1}) = k$ can be solved for $x^{n+1}$ as a function of $x^1, \dots, x^n$ near an initial point $\big(x^1(t_0), \dots, x^{n+1}(t_0)\big)$, and is equivalent to an equation $x^{n+1} = h(x^1, \dots, x^n)$ in some open set $U \subseteq \mathbb{R}^{n+1}$ containing the initial point $\big(x^1(t_0), \dots, x^{n+1}(t_0)\big)$. (That is, we are assuming $\partial f/\partial x^{n+1} \neq 0$.) To avoid trivial cases, we note that if $(x^j)'(t_0) = 0$ for all $j$, then the unique solution of this initial-value problem is the set of constant functions $x^j(t) = x^j(t_0)$. We shall therefore assume that $(x^1)'(t_0) \neq 0$ and hence $(x^1)'(t) \neq 0$ for $t$ near $t_0$. Again, if $M_j(x^1, \dots, x^{n+1}) = 0$ at the initial point, the corresponding expression isn't really defined. In that case, the corresponding variable can be assumed constant as well, and that equation can be eliminated from the system. From now on, we assume that none of the functions $M_j$ vanishes and that $(x^1)'(t) \neq 0$

We first show that the gradient of $f(x^1, \dots, x^{n+1})$ is orthogonal to the vector $\big(M_1(x^1, \dots, x^n), \dots, M_{n+1}(x^1, \dots, x^n)\big)$ at the points where $(x^1, \dots, x^{n+1}) = \big(x^1(t), \dots, x^{n+1}(t)\big)$ when $x^1(t), \dots, x^{n+1}(t)$ satisfy Eq. (7). To that end, we observe that

$$0 = \varphi'(t) = \frac{\partial f}{\partial x^1}(x^1)'(t) + \cdots + \frac{\partial f}{\partial x^{n+1}}(x^{n+1})'(t)\,.$$

If we multiply this equation by $M_1(x^1, \dots, x^{n+1})$ and exploit the equations

$$M_1(x^1, \dots, x^{n+1})x_j'(t) = M_j(x^1, \dots, x^{n+1})(x^1)'(t)\,,$$

we get the required relation immediately, since by assumption $(x^1)' \neq 0$:

$$M_1\frac{\partial f}{\partial x^1} + \cdots + M_{n+1}\frac{\partial f}{\partial x^{n+1}} \equiv 0 \tag{8}$$

at points $\big(x^1(t), \dots, x^{n+1}(t)\big)$ satisfying the differential equation and the conservation law $f\big(x^1(t), \dots, x^{n+1}(t)\big) = k$.

Since we are going to be replacing $x^{n+1}$ by $h(x^1, \dots, x^n)$ quite frequently, we need some notation to make our work easier. To that end, we introduce the operator $^*$ that maps a function $g(x^1, \dots, x^{n+1})$ to the function $g^*(x^1, \dots, x^n) = f\big(x^1, \dots, x^n, h(x^1, \dots, x^n)\big)$. Thus, we have $f^*(x^1, \dots, x^n) \equiv k$ when $x^1(t), \dots, x^{n+1}(t)$

satisfy the system of differential equations. We are going to apply this operation many times below, and so we note two easily proved but important properties of it: (1) If $\Phi(x^1,\dots,x^{n+1})$ is independent of $x^{n+1}$, then $\Phi^*(x^1,\dots,x^n) = \Phi(x^1,\dots,x^{n+1})$; (2) for any real-valued function of $m$ variables $\Psi(z^1,\dots,z^m)$, and any $m$ functions $\Phi_j(x^1,\dots,x^{n+1})$, $j=1,\dots,m$, the relation $\big(\Psi(\Phi_1,\dots,\Phi_m)\big)^* = \Psi(\Phi_1^*,\dots,\Phi_m^*)$ holds.

By the first of these, we have $h^* = h$ and $\partial h^*/\partial x^j = \partial h/\partial x^j$. Then the chain rule implies

$$\frac{\partial \Phi^*}{\partial x^j} = \Big(\frac{\partial \Phi}{\partial x^j}\Big)^* + \frac{\partial h}{\partial x^j}\Big(\frac{\partial \Phi}{\partial x^{n+1}}\Big)^* .$$

If we take $\Phi = f$ in this last relation, the fact that $\partial f^*/\partial x^j \equiv 0$ yields the relation

$$\Big(\frac{\partial f}{\partial x_j}\Big)^* = -\frac{\partial h}{\partial x^j}\Big(\frac{\partial f}{\partial x^{n+1}}\Big)^* ,$$

which in turn implies (if we divide Eq. (8) by $\partial f/\partial x^{n+1}$ and then apply the operator $*$) that

$$M_1^*\frac{\partial h}{\partial x^1} + \cdots + M_n^*\frac{\partial h}{\partial x^n} = M_{n+1}^* . \tag{9}$$

We now claim that the reduced system

$$\frac{dx^1}{M_1^*(x^1,\dots,x^n)} = \cdots = \frac{dx^n}{M_n^*(x^1,\dots,x^n)} , \tag{10}$$

will satisfy the condition that the divergence vanishes, provided it is multiplied by

$$\frac{1}{\mu(x^1,\dots,x^n)} = \Big(\frac{\partial f}{\partial x^{n+1}}\Big)^* .$$

In other words, we claim that

$$\frac{\partial}{\partial x^1}\left(\frac{M_1^*}{\big(\frac{\partial f}{\partial x^{n+1}}\big)^*}\right) + \cdots + \frac{\partial}{\partial x^n}\left(\frac{M_n^*}{\big(\frac{\partial f}{\partial x^{n+1}}\big)^*}\right) \equiv 0 .$$

The proof of this result is a routine computation, using Eqs. (8)–(9), and is left as an exercise (Problem 5.6). □

**Corollary 5.1.** *If $n-1$ independent integrals can be found for the divergence-free system* (7)*, it can be reduced to a single exact differental equation in two variables.*

Proof. The proof is a simple induction on $n$. The case $n=1$ is precisely that of an exact differential equation, for which nothing needs to be proved. The case $n = r+1$ reduces to the case $n = r$ by the theorem. □

## 4. An Application: Rotational Motion*

An excellent example of the use of the last-multiplier principle in classical physics is provided by the case of the motion of a rigid body about a fixed point. In relativistic mechanics, there are no rigid bodies, due to the observer-dependence of time at different points of the body, but in classical mechanics with its absolute simultaneity, there are such things. In order to present this example, we need first to discuss some vector calculus and then apply Newtonian mechanics using rotating coordinate systems.

Before we begin, this seems an appropriate place to make some remarks about vector spaces in general. Since, except for Chapter 3, this entire book is concerned with mechanics, there are only three physical "dimensions" that we need to be concerned with, namely mass ($m$), length—or distance—($s$), and time ($t$). As all of these are continuous quantities, and we use real numbers to express their magnitudes, we need to specify the units of each type. We are adhering to the MKS system, whereby the unit of length is the meter, the unit of mass is the kilogram, and the unit of time is the second. Other physically significant observable quantities are defined as products or quotients of these. Velocity ($v$) is the quotient when length is divided by time ($v = s/t$), momentum ($p$) is the product of mass and velocity ($p = mv = ms/t$), acceleration ($a$) is velocity divided by time ($a = v/t = s/t^2$), force $f$ is momentum divided by time ($f = p/t = mv/t = ms/t^2 = ma$), work $w$ is force times distance ($w = fs = ms^2/t^2 = mv^2$) and kinetic energy $e$ is mass times the square of a velocity, the same physical dimension as work. Angular momentum ($l$) is momentum times distance ($l = ps = ms^2/t$), and torque ($n$) is angular momentum divided by time ($n = l/t = ms^2/t^2$), once again, the same physical dimension as work or energy.

In order for these formal algebraic operations to be applicable to the observable world, the units in which each of the compound must be such that one unit of each type, when multiplied by the corresponding units of the other types result in one unit of the compound object. Thus, we express velocity in meters per second, so that an object that moves one meter in one second has unit velocity. Then an object that moves $x$ meters in $y$ seconds has a velocity of $x/y$ meters per second. Similarly, force is measured in newtons, one newton being the force required to impart to one kilogram of mass an acceleration of one meter per second per second. Thus the force required to impart an acceleration of $y$ meters per second per second to a mass of $x$ kilograms is $xy$ newtons, and so on.

Of the three basic quantities, two—mass and time—are measured as real numbers having no direction. In contrast, distances are measured between points having coordinates in a three-dimensional physical space $\mathbb{R}^3$ of positions. For that reason, we shall always assume that the components of a vector in $\mathbb{R}^3$ have the physical dimension of length. Most physical quantities that involve distance will also be vectors. Exceptions are work and energy, which do not depend on the individual components of the distance, but only on its absolute value. Now, in most elementary physics books, we find things that superficially appear to be absurd, such as drawings in which a vector representing a velocity is represented on a space of positions. We know that it is not really possible to add a velocity to a length. But we also understand the convenience of such a drawing, and we make it rigorous by keeping in mind that the velocity (or other quantity) really "inhabits" a second vector space attached as a tangent space to the manifold $\mathbb{R}^3$ at some point. The tangency means only that the derived quantity, whether force, angular momentum, or any other quantity, "inherits" its direction from the space of positions. As long as the units for each quantity are chosen as above, any two inertial observers (observers for whom Newton's first law holds) have only to establish when they are talking about the same location in physical space. We take for granted that they both measure mass and time the same way and get the same value. If they know the equations of transformation between the coordinates that they assign to points, they can use those same equations of transformation to reconcile the coordinates they assign to any other quantity whatsoever. All that is very simple, and

mostly unnecessary, since there is seldom any need to consider two different inertial systems.

But, if one system is inertial and the other isn't, as in the case of a rotating coordinate system in Newtonian absolute space, some terms of adjustment will have to be added to the basic coordinate transformation in order to reconcile each of the other quantities just listed. That is the first labor we need to undertake. We are going to assume that Observer 1 does observe Newton's first law, and so is using a right-handed orthonormal basis of position vectors $\{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$ to express the location of a given point $P$ as, say $x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$. But we shall not assume that Observer 2 has such a system. All we wish to insist on is that (1) Observer 2, like Observer 1, is using a right-handed orthonormal basis of position vectors to express the position of a point, and (2) the two observers agree as to the absolute value of the distance $d(P, Q)$ between any two points $P$ and $Q$. Thus, if Observer 1 assigns to $Q$ the coordinates $u\boldsymbol{i} + v\boldsymbol{j} + w\boldsymbol{k}$, then $d(P, Q) = \sqrt{(x-u)^2 + (y-v)^2 + (z-w)^2}$, and Observer 2 agrees that that is the distance. We first need to see what the coordinate transformation between the two has to be in order for that relation to be invariant between the two observers.

**4.1. Isometries.** In what follows, the vector quantity $\boldsymbol{u} \in \mathbb{R}^3$ is used to represent the coordinates that Observer 1 assigns to a given location in $n$-dimensional Euclidean space $\mathbb{R}^n$, and $T(\boldsymbol{u})$ is a vector representing the coordinates that Observer 2 assigns to the same point. We are assuming that $|T(\boldsymbol{u}) - T(\boldsymbol{v})| = |\boldsymbol{u} - \boldsymbol{v}|$ for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. It is obvious that such a function—called an *isometry*—is continuous. (In the "$\varepsilon$–$\delta$" technique, one can choose $\delta = \varepsilon$, no matter what $\varepsilon$ one is challenged with.) We intend to show that an isometry that leaves one point fixed is necessarily a rotation about that point or a rotation followed by a reflection.

To that end, we eliminate one simple case at the outset. Translation by a fixed vector $\boldsymbol{w} \in \mathbb{R}^n$ is an isometry, as it is trivial to show that if $T(\boldsymbol{u}) = \boldsymbol{w} + \boldsymbol{u}$, then $|T(\boldsymbol{u}) - T(\boldsymbol{v})| = |\boldsymbol{u} - \boldsymbol{v}|$. From now on, we shall assume that some point in the absolute Newtonian (Euclidean) space measured by Observer 1 also has constant coordinates at all times in the coordinates used by Observer 2. More specifically, we take this fixed point as the origin $\mathbf{0}$ of right-handed rectangular coordinate systems used by the two.

Under these circumstances, we claim that for any two points $\boldsymbol{u}$ and $\boldsymbol{v}$, the midpoint $\boldsymbol{w}$ of the line joining them—$\boldsymbol{w} = (\boldsymbol{u} + \boldsymbol{v})/2$—maps to the midpoint of the line joining $T(\boldsymbol{u})$ and $T(\boldsymbol{v})$, that is $T(\boldsymbol{w}) = \big(T(\boldsymbol{u}) + T(\boldsymbol{v})\big)/2$. To see this, let $r = |\boldsymbol{v} - \boldsymbol{u}|/2$. The closed $(n-1)$-dimensional spheres of radius $r$ about $\boldsymbol{u}$ and $\boldsymbol{v}$, that is, the points $\boldsymbol{z} \in \mathbb{R}^n$ such that $|\boldsymbol{z} - \boldsymbol{v}| = r$ and $|\boldsymbol{z} - \boldsymbol{u}| = r$ have in common only the one point $\boldsymbol{w} = (\boldsymbol{u} + \boldsymbol{v})/2$. This is a consequence of the triangle inequality in $\mathbb{R}^n$, which asserts that $|\boldsymbol{a} + \boldsymbol{b}| < |\boldsymbol{a}| + |\boldsymbol{b}|$ unless one of $\boldsymbol{a}$ and $\boldsymbol{b}$ is a nonnegative scalar multiple of the other. Thus, if $|\boldsymbol{z} - \boldsymbol{u}| = r = |\boldsymbol{v} - \boldsymbol{z}|$, the equality $|(\boldsymbol{z} - \boldsymbol{u}) + (\boldsymbol{v} - \boldsymbol{z})| = |\boldsymbol{v} - \boldsymbol{u}| = 2r$ implies that $\boldsymbol{z} - \boldsymbol{u} = t(\boldsymbol{v} - \boldsymbol{z})$ for $t > 0$. That equation implies $r = |\boldsymbol{z} - \boldsymbol{u}| = t|\boldsymbol{v} - \boldsymbol{z}| = tr$, and so $t = 1$, that is, $\boldsymbol{z} = (\boldsymbol{u} + \boldsymbol{v})/2$. But since the mapping $T$ preserves distances, we also have $|T(\boldsymbol{z}) - T(\boldsymbol{u})| = r = |T(\boldsymbol{u}) - T(\boldsymbol{z})|$, and $\big|\big(T(\boldsymbol{z}) - T(\boldsymbol{u})\big) + \big(T(\boldsymbol{v}) - T(\boldsymbol{z})\big)\big| = 2r$, and so $T(\boldsymbol{z}) - T(\boldsymbol{u}) = t\big(T(\boldsymbol{v}) - T(\boldsymbol{u})\big)$ for some $t > 0$, which again implies that $T(\boldsymbol{z}) = \big(T(\boldsymbol{u}) + T(\boldsymbol{v})\big)/2$.

Applying this same reasoning to the pairs $\boldsymbol{u}, (\boldsymbol{u}+\boldsymbol{v})/2$ and $(\boldsymbol{u}+\boldsymbol{v})/2, \boldsymbol{v}$, we see that it is also true that

$$T\Big(\frac{1}{4}\boldsymbol{u}+\frac{3}{4}\boldsymbol{v}\Big) = \frac{1}{4}T(\boldsymbol{u}) + \frac{3}{4}T(\boldsymbol{v}) \text{ and } T\Big(\frac{3}{4}\boldsymbol{u}+\frac{1}{4}\boldsymbol{v})\Big) = \frac{3}{4}T(\boldsymbol{u}) + \frac{1}{4}T(\boldsymbol{v})\,.$$

Continuing in this way, we eventually see that for every binary rational number $c = r/2^s$ between 0 and 1 ($0 \le r \le 2^s$), we have

$$T\big((1-c)\boldsymbol{u}+c\boldsymbol{v}\big) = (1-c)T(\boldsymbol{u}) + cT(\boldsymbol{v})\,.$$

Since, for fixed values of $\boldsymbol{u}$ and $\boldsymbol{v}$ the functions of $c$ on the two sides of this last equation are both continuous, and they concide on the dense set of binary rationals in $[0,1]$, it follows that this relation holds for all real scalars $c \in [0,1]$. But now, if $c > 1$ and $\boldsymbol{w} = (1-c)\boldsymbol{u} + c\boldsymbol{v}$, we see that $\boldsymbol{v} = \frac{1}{c}\boldsymbol{w} + \big(1-\frac{1}{c}\big)\boldsymbol{u}$ and $0 < \frac{1}{c} < 1$, so that we get the relation $T(\boldsymbol{v}) = \frac{1}{c}T(\boldsymbol{w}) + \big(1-\frac{1}{c}\big)T(\boldsymbol{u})$, which then translates to $T\big((1-c)\boldsymbol{u}+c\boldsymbol{v}\big) = T(\boldsymbol{w}) = (1-c)T(\boldsymbol{u}) + cT(\boldsymbol{v})$. Thus, the relation holds for arbitrary positive values of $c$. In particular, taking $\boldsymbol{u} = \boldsymbol{0}$, we have $T(c\boldsymbol{u}) = cT(\boldsymbol{u})$ for all positive values of $c$. Taking $c = \frac{1}{2}$ and $\boldsymbol{v} = -\boldsymbol{u}$, we get $\boldsymbol{0} = T(\boldsymbol{0}) = \frac{1}{2}\big(T(\boldsymbol{u}) + T(-\boldsymbol{u})\big)$, so that $T(-\boldsymbol{u}) = -T(\boldsymbol{u})$. It then follows that $T(c\boldsymbol{u}) = cT(\boldsymbol{u})$ even for negative values of $c$, as we see by replacing $c$ with $-c$ and $\boldsymbol{u}$ with $-\boldsymbol{u}$ when $c < 0$. Thus, the formula now holds in general, and we have in particular $T\big(\boldsymbol{u}+\boldsymbol{v}\big) = T\big(2(\boldsymbol{u}/2+\boldsymbol{v}/2)\big) = 2T\big((\boldsymbol{u}+\boldsymbol{v})/2\big) = 2T(\boldsymbol{u}/2) + 2T(\boldsymbol{v}/2) = T(\boldsymbol{u}) + T(\boldsymbol{v})$. It follows that $T$ preserves scalar multiplication and vector addition. That is, we have proved that the only transformations that have a fixed point and preserve distance are linear transformations.

They are better than just linear, however. They also preserve the dot product, which is to say, the angle between two vectors. Indeed, by the polarization formula, we have

$$T(\boldsymbol{u})\cdot T(\boldsymbol{v}) = \frac{\big|T(\boldsymbol{u})+T(\boldsymbol{v})\big|^2}{4} + \frac{\big|T(\boldsymbol{u})-T(\boldsymbol{v})\big|^2}{4} = \frac{|\boldsymbol{u}+\boldsymbol{v}|^2}{4} + \frac{\big|\boldsymbol{u}-\boldsymbol{v}\big|^2}{4} = \boldsymbol{u}\cdot\boldsymbol{v}\,.$$

This means that the adjoint $T^\tau$, which is defined by the equation $T(\boldsymbol{u})\cdot\boldsymbol{v} = \boldsymbol{u}\cdot T^\tau(\boldsymbol{v})$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$, is the inverse of $T$, that is $T^\tau T = I$, the identity transformation on $\mathbb{R}^n$. For since $\boldsymbol{u}\cdot\boldsymbol{v} = T(\boldsymbol{u})\cdot T(\boldsymbol{v}) = (\boldsymbol{u})\cdot T^\tau\big(T(\boldsymbol{v})\big)$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$, it follows that $T^\tau\big(T(\boldsymbol{v})\big) = \boldsymbol{v}$ for all $\boldsymbol{v}$. It is well-known from linear algebra that the matrix of $T^\tau$, in any orthonormal basis, is the transpose of the matrix of $T$, obtained by interchanging rows and columns. In particular, if $\boldsymbol{u}_i$ is the $i$th row and $\boldsymbol{v}_j$ is the $j$th row, then $\boldsymbol{u}_i\cdot\boldsymbol{v}_j$ is the $i,j$ entry of the matrix of the product $TT^\tau$, and hence equals the Kronecker delta $\delta_i^j$ (see Chapter 5 of Volume 1), which is 1 if $i = j$ and 0 if $i \ne j$. In particular, the rows of the matrix form an orthonormal basis of $\mathbb{R}^n$, as do the columns. A matrix with this property is said to be *orthogonal*.

**4.2. Rotations.** The fact that an orthogonal transformation $T$ preserves lengths and angles means that it maps the unit cube whose corners are at the $2^n$ points $(\varepsilon_1, \dots, \varepsilon_n)$, where $\varepsilon_j = 1$ or 0 for each $j$, to another unit cube (all its right angles and side lengths are preserved). Hence the $n$-dimensional volume of this cube does not change. But the volume of the image of this unit cube under $T$ is the absolute value of the determinant of the transformation $T$, that is, of its matrix in any orthonormal basis. In short, the determinant of $T$ is $\pm 1$. It also follows that if $\boldsymbol{u}$ is an eigenvector of $T$, then $T(\boldsymbol{u}) = \pm\boldsymbol{u}$, since $T(\boldsymbol{u}) = c\boldsymbol{u}$ for some scalar (by definition of an eigenvector) and $|T(\boldsymbol{u})| = |c\boldsymbol{u}| = |c|\,|\boldsymbol{u}| = |c||T(\boldsymbol{u})|$. Thus, we reach

two conclusions: (1) the only real eigenvalues $T$ can have are the two numbers $\pm 1$, and (2) the determinant of the transformation, which is the product of all the eigenvalues (real or complex) is also $\pm 1$.

To specialize these considerations to the low-dimensional cases, first consider $n = 2$. The characteristic polynomial has either two real roots (eigenvalues), which must be $\pm 1$, or it has a pair of conjugate complex roots whose product is $\pm 1$. Since the product of two conjugate complex numbers is non-negative, in the second case, this product must be 1, and each of the eigenvalues has absolute value 1. That means that the eigenvalues are of the form $\cos\theta \pm i \sin\theta$ for some angle $\theta$. Since the unique monic polynomial having these roots is $\lambda^2 - 2\cos\theta\lambda + 1$, it follows that the trace of $T$ in this case is $2\cos\theta$. One can easily deduce that, by a suitable choice of an orthonormal basis of $\mathbb{R}^2$, it is possible to arrange that the matrix of $T$ shall be

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

This is easy to do, since $r_{11}^2 + r_{12}^2 = 1$, so that there definitely exists an angle $\theta$ such that $r_{11} = \cos\theta$ and $r_{12} = \sin\theta$. Since $r_{11}^2 + r_{21}^2 = 1 = r_{12}^2 + r_{22}^2$, it follows that $r_{21} = \varepsilon_1 \sin\theta$ and $r_{22} = \varepsilon_2 \cos\theta$, where $\varepsilon_1 = \pm 1$ and $\varepsilon_2 = \pm 1$. The determinant is therefore $\varepsilon_2 \cos^2\theta - \varepsilon_1 \sin^2\theta$. Since the determinant is to be $+1$—in which case the transformation is called a *rotation*— we need $\varepsilon_1 = -1$ and $\varepsilon_2 = +1$, as asserted.

Thus, when there are no real eigenvalues, the transformation is necessarily a rotation. It can also be a trivial rotation in the case when the eigenvalues are both $+1$ ($T$ is the identity transformation $I$, which is a "rotation" through 0 degrees) or both $-1$ ($T$ is $-I$, the negative of the identity transformation, which is a rotation through 180 degrees). Only when one eigenvalue is $+1$ and the other $-1$ does the mapping $T$ fail to be a rotation. In this case, it is a reflection, which can be regarded as a rotation by 180 degrees about some fixed line in the plane when the plane is embedded in $\mathbb{R}^3$. But this transformation cannot be reached as the end result of a continuous family of rigid motions $T(t)$ starting from $T(0) = I$. That is because the determinant is a continuous function of $t$ taking only only the values $\pm 1$, and therefore required to be constant for all $t$. Thus, in terms of physical motion of a rigid body, starting from rest, the only possible motions are rotations. And, of course, this same result holds (for the same reason) in all Euclidean spaces $\mathbb{R}^n$. Since we shall be dealing only with rotations from this point on, we shall now use the symbol $R(t)$ to denote the transformation that takes the "rotating" coordinates of a point $P$ in space (those being used by Observer 2) to the coordinates of the location of the same point $P$ in the system used by the inertial observer. We may assume without any loss of generality that $R(0)$ is the trivial identity rotation, in other words, that the inertial and rotating coordinate axes coincide at the initial time. Then, by what was proved above, the "rotating" coordinates of a point are obtained from its inertial coordinates by the inverse/adjoint operation $R^\tau(t)$, whose matrix is the transpose of the matrix of $R(t)$. In short, the entries in the $j$th column of $R(t)$ are the coordinates assigned by Observer 1 to the $j$th basis vector in the orthonormal system used by Observer 2. The $i$th row of this matrix (the $i$th column of its inverse $R^\tau(t)$) contains the coordinates that Observer 2 assigns to the $i$th basis vector in the orthonormal system used by Observer 1. If Observer 1 is using the orthonormal basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ and Observer 2 the orthonormal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$, then $r_{ij} = \boldsymbol{u}_i \cdot \boldsymbol{v}_j$, which, since both vectors have unit length, is the cosine of the angle between the $i$th inertial axis and the $j$th rotating axis. These

entries are therefore called the *direction cosines* of the rotating axes relative to the inertial axes.

The most important special case is the case of $\mathbb{R}^3$. In this space a rotation preserves the not only the dot product, but also the cross product. This is because the cross product $\boldsymbol{u}\times\boldsymbol{v}$ is characterized by the facts that (1) its length is $|\boldsymbol{u}|\,|\boldsymbol{v}|\sin\theta$, where $\theta$ is the angle between the two vectors, (2) its direction is perpendicular to the plane spanned by $\boldsymbol{u}$ and $\boldsymbol{v}$ and such that $\{\boldsymbol{u},\boldsymbol{v},\boldsymbol{u}\times\boldsymbol{v}\}$ is a right-handed system. A rotation preserves all these things because it preserves lengths, angles, and right-handed systems. Thus, from now on $R(t)$ is represented by a $3\times 3$ rotation matrix varying continuously as a function of $t$.

A basic problem of classical mechanics is to describe the motion of a rigid body with a fixed points by using Newton's laws to get a formula for $R(t)$. The relation $R^\tau(t)R(t) = I$ provides a set of six equations for the nine entries in the matrix $R(t) = \{r_{ij}(t)\}_{i,j=1,2,3}$, namely

$$\sum_{k=1}^{3} r_{ki}(t)r_{kj}(t) = \delta_i^j \quad 1\le i\le j\le 3\,. \tag{11}$$

These six relations leave $R(t)$ with only three degrees of freedom,[5] that is, arbitrariness that we would like to eliminate so as to determine $R(t)$ uniquely. We need three more equations to determine the functions $r_{ij}(t)$. We get them by differentiating the six that we already have. Nothing new comes of differentiating the equations in (11) with $i = j$, but if $i \neq j$, we do get three new quantities, namely

$$\begin{aligned}
\omega_1(t) &= \sum_{k=1}^{3} r_{k3}(t)r'_{k2}(t) = -\sum_{k=1}^{3} r'_{k3}(t)r_{k2}(t)\,,\\
\omega_2(t) &= \sum_{k=1}^{3} r_{k1}(t)r'_{k3}(t) = -\sum_{k=1}^{3} r'_{k1}(t)r_{k3}(t)\,,\\
\omega_3(t) &= \sum_{k=1}^{3} r_{k2}(t)r'_{k1}(t) = -\sum_{k=1}^{3} r'_{k2}(t)r_{k1}(t)\,.
\end{aligned}$$

These three differential equations, together with the values of the entries in $R(0)$, would enable us to find all nine of the entries $r_{ij}(t)$, provided we knew the value of $\boldsymbol{\omega}(t) = \omega_1(t)\boldsymbol{i} + \omega_2(t)\boldsymbol{j} + \omega_3(t)\boldsymbol{k}$. That is the quantity we shall be looking for. We are interested in this problem only in application to the motion of a rotating rigid body $B$ relative to which the coordinates used by Observer 2 are fixed. As we shall eventually see, if $\boldsymbol{N}(t)$ are the components of the torque on the body $B$ at time $t$, as measured by the inertial observer, and $\boldsymbol{\Lambda}(t) = R^\tau(t)\big(\boldsymbol{N}(t)\big)$, we can get a system of first-order ordinary differential equations in the two vector-valued functions $\boldsymbol{\omega}(t)$ and $\boldsymbol{\Lambda}(t)$, which amounts to a system of six differential equations in the components of these vectors. Given initial values then, $\boldsymbol{\omega}(t)$ and $\boldsymbol{\Lambda}(t)$ are both determined theoretically, whether or not one can find their values computationally. Having $\boldsymbol{\omega}(t)$, we can then forget about $\boldsymbol{\Lambda}(t)$ and solve the system of equations above to get the values of all nine entries in $R(t)$ at any time.

This discussion of rotations is meant as a lead-in to establishing the transformations by which the two observers observers reconcile their measurements of

[5] Putting this another way, the group of rotations of $\mathbb{R}^3$ is a three-dimensional Lie group.

important vector physical quantites: velocity, momentum, angular momentum, acceleration, force, and torque.

**4.3. Dimension and direction.** In the following discussion, we will be mentioning many vector spaces, each of which will have two sets of orthonormal coordinates, one for an inertial observer privileged to use Newton's laws in the simplest vector form, and another fixed in a rotating rigid body, for which Newton's laws still apply, but only if certain extra forces are acknowledged. What all of these have in common is (1) an origin whose coordinates are $(0,0,0)$ in both systems, and (2) a basic set of mutually orthogonal directions, all determined by the directions in the fundamental space of positions. Those positions are given as distances along the three axes from the origin, and hence all coordinates in the fundamental spaces have the physical dimension of length. The directions in all the spaces derived from the fundamental space have physical dimensions of velocity, acceleration, momentum, force, angular velocity, angular acceleration, angular momentum, and torque. Each of these derived spaces is defined by performing certain operations (scalar multiplication or cross product or differentiation) on vectors denoting the position of a physical point, and that is how the directions are inherited from the fundamental space. The rotation operators just discussed are dimensionless and hence can be applied at will to vectors in any and all of these spaces. Our first job is to see how the coordinates in an inertial system are reconciled with those in a rotating system, for all of these vector quantities.

**4.4. Relations between inertial and rotating coordinates.** Let us first consider velocity. Let $R(t)$ be the rotation matrix such that $\boldsymbol{u}(t) = R(t)\big(\boldsymbol{v}(t)\big)$ is the set of inertial coordinates at time $t$ of the point whose coordinates in the rotating system are $\boldsymbol{v}(t)$. It is obvious that the coordinates of the velocity of the particle that is at $\boldsymbol{u}(t)$ at time $t$ must be $\boldsymbol{u}'(t) = R(t)\big(\boldsymbol{v}'(t)\big) + R'(t)\big(v(t)\big)$. In other words, the conversion $R(t)\big(\boldsymbol{v}'(t)\big)$, which is all that would be needed between two inertial observers (since $R(t)$ would be constant), must be corrected by the addition of the term $R'(t)\big(v(t)\big)$, which depends on the coordinates of position. (No correction is needed, for example, for the point whose coordinates are $(0,0,0)$. These are the same in both systems.) Let us apply some linear algebra to this expression.

Considering that $R^\tau(t)R(t) = I$ (the identity matrix) and the obvious fact that for any two matrix-valued functions of time, $A(t)$ and $B(t)$, we have $\big(A(t)B(t)\big)' = A'(t)B(t) + A(t)B'(t)$, together with the equally obvious facts that $(R^\tau)'(t) = \big(R'(t)\big)^\tau$, $(AB)^\tau = B^\tau A^\tau$, and $\big(A^\tau\big)^\tau = A$, we easily find that

$$\boldsymbol{O} = (R^\tau)'(t)R(t) + R^\tau(t)R'(t) = \big(R'(t)\big)^\tau R(t) + R^\tau(t)R'(t) = \\ = \big(R^\tau(t)R'(t)\big)^\tau + R^\tau(t)R'(t)\,,$$

where $\boldsymbol{O}$ is the $3\times 3$ matrix whose entries are all zeros, so that

$$(R^\tau R')^\tau = -R^\tau R'\,.$$

This equation means that the matrix $R^\tau R'$ is *skew-symmetric* and is in fact

$$R^\tau R' = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}\,,$$

where $\boldsymbol{\omega} = \omega_1\boldsymbol{i} + \omega_2\boldsymbol{j} + \omega_3\boldsymbol{k}$ is the vector introduced above. Its coordinates in the rotating system are $(\omega_1, \omega_2, \omega_3)$, and it represents the angular velocity of the rotating system relative to the inertial system, in which we shall give it the coordinates $(w_1, w_2, w_3)$. These transform into each other straightforwardly: $\big(w_1(t), w_2(t), w_3(t)\big) = R(t)\big(\omega_1(t), \omega_2(t), \omega_3(t)\big)$. It should be obvious from this relation that $|\boldsymbol{\omega}| = |\boldsymbol{w}|$, even if it was not already obvious on the grounds that angles have an absolute meaning and both observers are using the common Newtonian time.

We chose this notation because it implies that if $\boldsymbol{v} = v_1\boldsymbol{i} + v_2\boldsymbol{j} + v_3\boldsymbol{k}$, then

$$R^\tau R'(\boldsymbol{v}) = (\omega_2 v_3 - \omega_3 v_2)\boldsymbol{i} + (\omega_3 v_1 - \omega_1 v_3)\boldsymbol{j} + (\omega_1 v_2 - \omega_2 v_1)\boldsymbol{k} = \boldsymbol{\omega} \times \boldsymbol{v}\,.$$

Putting all this together, we get the equation

$$R^\tau(t)\big(\boldsymbol{u}'(t)\big) = \boldsymbol{v}'(t) + \boldsymbol{\omega}(t) \times \boldsymbol{v}(t)\,.$$

This relation is trivially equivalent to

$$\boldsymbol{v}'(t) = R^\tau(t)\big(\boldsymbol{u}'(t)\big) - \boldsymbol{\omega}(t) \times \boldsymbol{v}(t)\,.$$

Since $R(t)$ preserves the cross product, we get the equivalent relation

$$\begin{aligned} \boldsymbol{u}'(t) &= R(t)\big(\boldsymbol{v}'(t)\big) + R(t)\big(\boldsymbol{\omega}(t) \times \boldsymbol{v}(t)\big) \\ &= R(t)\big(\boldsymbol{v}'(t)\big) + R(t)\big(\boldsymbol{\omega}(t)\big) \times R(t)\big(\boldsymbol{v}(t)\big) \\ &= R(t)\big(\boldsymbol{v}'(t)\big) + \boldsymbol{w}(t) \times \boldsymbol{u}(t)\,. \end{aligned}$$

We repeat that $\boldsymbol{w}(t) = R(t)\big(\boldsymbol{\omega}(t)\big)$, where $\boldsymbol{\omega}(t)$ is defined as above, gives the inertial coordinates of the *angular velocity* of the rotating coordinate system relative to the inertial system. While the cross product can be applied to two vectors of different physical dimension, it is necessary for the coordinates of the two factors to "belong" to the same basic system of axes. Therefore the vector $\boldsymbol{\omega}(t)$ is the set of coordinates that the rotating system assigns to its own rotation relative to the inertial system. The negatives of these two quantities are the angular velocities that the two assign to the inertial system relative to the rotating system.

The reasoning that established the relation $\boldsymbol{u}'(t) = R(t)\big(\boldsymbol{v}'(t)\big) + \boldsymbol{w}(t) \times \boldsymbol{u}(t)$ is quite general and applies when $\boldsymbol{v}$ is replaced by $\boldsymbol{\omega}$ and $\boldsymbol{u}$ by $\boldsymbol{w}$. Thus, we find

$$\boldsymbol{w}'(t) = R(t)\big(\boldsymbol{\omega}'(t)\big) + \boldsymbol{w} \times \boldsymbol{w}(t) = R(t)\big(\boldsymbol{\omega}'(t)\big)\,.$$

In other words, the time derivatives of angular velocity also transform just like position vectors.

These formulas tell us how two observers know when they are observing the same velocity for a particle. In the case when the rotating coordinates behave like a "sleeping top," $\boldsymbol{\omega}(t)$ and $\boldsymbol{w}(t)$ are both constant. Then, from the inertial point of view, a particle having constant rotating coordinates $\boldsymbol{v}$ at all times and inertial coordinates $\boldsymbol{u}(t)$ at time $t$ is rotating about the line through the fixed point in the direction of $\boldsymbol{w}$ at constant linear speed $v = |\boldsymbol{u}'(t)| = |\boldsymbol{\omega} \times \boldsymbol{u}(t)|$. Its distance from that line is $r = |\boldsymbol{u}(t)| \sin\theta = |\boldsymbol{u}| \sin\theta = |\boldsymbol{w} \times \boldsymbol{u}|/|\boldsymbol{w}| = v/|\boldsymbol{w}|$, where $\theta$ is the angle between $\boldsymbol{u}$ and $\boldsymbol{w}$. This particle is moving around a circle of radius $r\sin\theta$ at constant linear speed $v$ equal to $|\boldsymbol{u}'(t)| = r|\boldsymbol{w}|\sin\theta$. Thus, not only does $\boldsymbol{w}$ point along the instantaneous axis of rotation, its length gives the angular speed of rotation in radians per unit time.

**4.5. Acceleration and force.** We now resume our study of the conversion of physical data between the two systems. Observer 1 is using a system fixed in absolute Newtonian space, and Observer 2 a system that is rotating in an arbitrary manner with time. Newton's laws (in vector form) hold for the first observer, and we need to see what laws they imply for the second observer. Thus, we picture a particle moving around in such a way that at time $t$ the coordinates $\boldsymbol{v}(t)$ assigned to its position by Observer 2, correspond to the coordinates $\boldsymbol{u}(t) = R(t)\boldsymbol{v}(t)$ assigned by Observer 1. As we know, $R^\tau(t)$ converts coordinates in the opposite direction.

To reconcile the components of the acceleration of a particle, suppressing the argument $t$, we find that

$$\begin{aligned} \boldsymbol{u}'' &= R(\boldsymbol{v}'') + R'(\boldsymbol{v}') + \boldsymbol{w}' \times \boldsymbol{u} + \boldsymbol{w} \times \boldsymbol{u}' \,, \\ R^\tau(\boldsymbol{u}'') &= \boldsymbol{v}'' + \boldsymbol{\omega} \times \boldsymbol{v}' + \boldsymbol{\omega}' \times \boldsymbol{v} + \boldsymbol{\omega} \times (\boldsymbol{v}' + \boldsymbol{\omega} \times \boldsymbol{v}) \\ &= \boldsymbol{v}'' + 2\boldsymbol{\omega} \times \boldsymbol{v}' + \boldsymbol{\omega}' \times \boldsymbol{v} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{v}) \,, \end{aligned}$$

so that

$$\begin{aligned} \boldsymbol{u}'' &= R(\boldsymbol{v}'') + \boldsymbol{w}' \times \boldsymbol{u} + 2\boldsymbol{w} \times (\boldsymbol{u}' - \boldsymbol{w} \times \boldsymbol{u}) + \boldsymbol{w} \times (\boldsymbol{w} \times \boldsymbol{u}) \\ &= R(\boldsymbol{v}'') + \boldsymbol{w}' \times \boldsymbol{u} + 2\boldsymbol{w} \times \boldsymbol{u}' - \boldsymbol{w} \times (\boldsymbol{w} \times \boldsymbol{u}) \,; \\ \boldsymbol{v}'' &= R^\tau(\boldsymbol{u}'') - \boldsymbol{\omega}' \times \boldsymbol{v} - 2\boldsymbol{\omega} \times \boldsymbol{v}' - \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{v}) \,. \end{aligned}$$

Multiplying these equations by the mass $m$ of a particle whose position at time $t$ is $\boldsymbol{u}(t)$ or $\boldsymbol{v}(t)$ in the two systems, we get the force coordinates $\boldsymbol{F}_i$ and $\boldsymbol{F}_r$ that Observers 1 and 2 believe are acting on the particle. As we have characterized them, Observer 1, using the inertial system, sees the situation with Newtonian clarity: the actual force on the particle is $m\boldsymbol{u}''$ in inertial coordinates. Observer 2 should know that the true force acting on the particle is $R^\tau\big(\boldsymbol{F}_i\big)$. But the force actually measured by Observer 2 is $m\boldsymbol{v}''$, and it differs from the true force by the three extra terms arising from this last formula. These terms can be regarded as three "fictitious" forces that Observer 2 must adjoin to the forces reported by Observer 1 (that is, the "true" forces) in order to use Newtonian mechanics.[6] The first of these forces is $-m\boldsymbol{\omega}' \times \boldsymbol{v}$: Observer 2 must adjust the "true" acceleration of the particle measured by Observer 1 by taking account of the angular acceleration of the rotating coordinate system he is using. Subtracting this term achieves that adjustment.

The second term $-2m\boldsymbol{\omega} \times \boldsymbol{v}'$ acts on a particle that Observer 2 regards as being in motion with velocity $\boldsymbol{v}'$, causing it to veer off from the direction of motion in a direction perpendicular to the plane of the velocity and the angular velocity. It is known as the *Coriolis force*[7] It is this force that causes rockets launched northward in the northern hemisphere to veer to the east. (It is intuitively obvious

[6] There is a subtlety here. Observer 2 will actually observe the effects of these three forces, and to that extent, Newton's laws demand that they be regarded as real forces in the rotating coordinate system. Presumably, Observer 1 has a non-tautological definition of the coordinates of the force $\boldsymbol{F}$ acting on the particle—the law of gravity, for example, or some electric or magnetic attraction or repulsion—and these coordinates necessarily agree with those of $m\boldsymbol{u}''$, and so transform into $R^\tau(\boldsymbol{u}'')$ when converted to the rotating system. But since that expression is *not* the same as $\boldsymbol{v}''$, Observer 2 must include the three extra forces in order to work within Newtonian mechanics.

[7] Named for Gaspard Gustave de Coriolis (1792–1843), a French engineer and mathematical scientist, who introduced it in his 1835 paper "Sur les équations du mouvement relatif des systèmes de corps." It was precisely the addition of this term to the equations of motion that made Newtonian analysis possible in rotating coordinate systems. Coriolis ought to be remembered for more than this. He played an important role in clarifying the definition of work and kinetic energy.

that they must do so, since they have a greater linear eastward velocity when launched than the points on earth at the higher latitudes they will be flying over.) It also explains why air masses in the northern hemisphere that expand outward from a high pressure region tend to have a clockwise circulation when viewed from above, while those in the southern hemisphere circulate counterclockwise around a region of high pressure.

The third term is an adjustment that depends on the position of the particle. If a point $\boldsymbol{v}$ is fixed in Observer 2's frame of reference, it is, in celestial terms, rotating about the instantaneous axis of rotation at speed $r|\boldsymbol{w}|$, where $r = |\boldsymbol{v}| \sin\theta$, $\theta$ being the angle between $\boldsymbol{v}$ and the instantaneous axis of rotation. It is therefore subject to a centripetal force $mr|\boldsymbol{w}|^2 = m(|\boldsymbol{v}| \sin\theta)|\boldsymbol{\omega}|^2$. Since Observer 2 measures no force on the particle, that observer must correct his list of forces by subtracting this centripetal force. In other words, he needs to add a constant centrifugal force equal to $-m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{v})$.

The first of these three adjustments is certainly to be expected: If there is acceleration in one's coordinate axes, then surely the acceleration of a particle measured using those axes will be affected. The presence of this term in the equation is thus not at all surprising. The other two "fictitious" forces—which are fictitious only in the sense that the privileged inertial observer does not measure them—require some explanation.

**Example 5.2.** To illustrate the Coriolis and centrifugal forces, we may consider that the center of the earth (or any other location in the solar system that is fixed in the sun or some planet or moon) is stationary relative to the stars. And since the earth's north-south axis points constantly toward the star Polaris,[8] we can fix an inertial coordinate system with origin at the center of the earth, one axis in the north-south direction, and two other axes at right angles to that one and to each other, also pointing at fixed stars, and ordered in such a way that the unit vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ along the first and second horizontal axes respectively and a unit vector $\boldsymbol{k}$ pointing toward the north pole form a right-handed system.

On the earth, we take the same north-south axis, and in the plane of the equator, we take two arbitrary mutually perpendicular axes, again ordered and oriented in such a way that unit vectors pointing in the positive direction along each of them form a right-handed system when a unit north-pointing vector is taken as the third basis vector for a rotating coordinate system. As the horizontal axes rotate at a uniform rate of $\omega = 7.29 \times 10^{-5}$ radians per second (one revolution per sidereal day, consisting of 86,164.1 seconds), we can measure time from any instant at which the body and spatial axes coincide. In that case, since the earth rotates from west to east, the rotation matrix $R(t)$ that converts terrestrial (rotating) coordinates to spatial (inertial) coordinates is

$$R(t) = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) & 0 \\ \sin(\omega t) & \cos(\omega t) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(This expression means that the rotating coordinate system is behaving like the "sleeping top" mentioned above.)

[8] Of course, because of precession, this axis is not *quite* constant in direction, but the precession takes so long—some 25,000 years for one revolution—that for day-to-day observations on the earth it may be regarded as constant.

The inertial coordinates of the point that initially had coordinates $(1,0,0)$ in both systems (which remain constant in the rotating system) will be $\big(\cos(\omega t),\sin(\omega t),0\big)$ after $t$ seconds. It is then trivially computable that

$$R^\tau(t)R'(t) = \begin{pmatrix} 0 & -\omega & 0 \\ \omega & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

which is to say, the angular velocity of the earth's rotation in absolute space corresponds to the constant vector $\boldsymbol{\omega} = \omega\boldsymbol{k}$, and $\boldsymbol{w} = \omega\boldsymbol{k}$ for the same reason. This fact allows us to clear the terms involving $\boldsymbol{\omega}'$ and $\boldsymbol{w}'$ from our equation and concentrate on the other two corrective terms.

If a moving point $\boldsymbol{v}(t)$ has spherical coordinates $\rho(t)$ (altitude above the center of the earth), $\varphi(t)$ (latitude), and $\theta(t)$ (longitude) when measured in terrestrial (rotating) coordinates, one can easily compute that its location in celestial (inertial) coordinates is

$$\boldsymbol{u}(t) = \rho(t)\cos\big(\varphi(t)\big)\cos\big(\theta(t)+\omega t\big)\,\boldsymbol{i}+\rho(t)\cos\big(\varphi(t)\big)\sin\big(\theta(t)+\omega t\big)\,\boldsymbol{j}+\rho\sin\big(\varphi(t)\big)\,\boldsymbol{k}\,.$$

Thus, converting from terrestrial coordinates to celestial coordinates is simply a matter of increasing the longitude by the amount $\omega t$ by which the earth has rotated between time $t$ and time 0, leaving the altitude and latitude unchanged. And, obviously, converting in the other direction requires only subtracting that same amount of longitude. Working now purely in terrestrial coordinates and anticipating the need to take the cross product $\boldsymbol{\omega}\times\boldsymbol{v}$, we observe that this velocity vector is the angular speed $\omega\cos\varphi$ times the *position* vector $\boldsymbol{v}_1$ whose altitude is $\rho$, whose latitude is $0°$, and whose longitude is $\theta+\pi/2$. In particular, $\boldsymbol{\omega}\times\boldsymbol{v}$ is zero at the poles, where $\varphi = \pm 90°$, and has its largest magnitude at locations in the equatorial plane, where $\varphi = 0$, the vector having magnitude $\rho\omega$ and pointing directly east at each such point. At latitudes intermediate between those of the equator and a pole, it points tangentially eastward on the sphere about the origin through that point. That is no surprise, of course, since, in addition to whatever velocity a particle at that point has in terrestrial coordinates, in celestial coordinates, that velocity is augmented by a uniform circular motion eastward at angular speed $\omega$ about the north-south axis in a plane perpendicular to the north-south axis and at a radius of $\rho\cos(\varphi)$. In those coordinates, its eastward motion is at speed $v = \omega\rho\cos(\varphi) = |\boldsymbol{\omega}\times\boldsymbol{v}|$. (The angle between $\boldsymbol{\omega}$ and $\boldsymbol{v}$ is $\pi/2-\varphi$, and hence its sine is $\cos(\varphi)$.)

We note that the velocity vector $\boldsymbol{\omega}(t)\times\boldsymbol{v}_1(t)$ is the angular speed $\omega(t)$ times the position vector $\boldsymbol{\psi}(t)$ given by

$$\boldsymbol{\psi}(t) = -\rho(t)\cos\big(\theta(t)\big)\boldsymbol{i} - \rho(t)\sin\big(\theta(t)\big)\boldsymbol{j}\,.$$

The conversion of the coordinates of the velocity vectors is also easy to compute:

$$\boldsymbol{v}'(t) = \frac{\rho'(t)}{\rho(t)}\boldsymbol{v}(t)\,\boldsymbol{i} + \cos\big(\varphi(t)\big)\theta'(t)\boldsymbol{v}_1(t)\,\boldsymbol{j} + \varphi'(t)\boldsymbol{v}_2(t)\,\boldsymbol{k}\,,$$

where $\boldsymbol{v}_1(t)$ is the position vector described above and $\boldsymbol{v}_2(t)$ is given by

$$\boldsymbol{v}_2(t) = -\rho(t)\sin\big(\varphi(t)\big)\cos\big(\theta(t)\big)\,\boldsymbol{i} - \rho(t)\sin\big(\varphi(t)\big)\sin\big(\theta(t)\big)\,\boldsymbol{j} + \rho(t)\cos\big(\varphi(t)\big)\,\boldsymbol{k}\,.$$

Under these circumstances, the Coriolis force per unit mass is

$$-2\boldsymbol{\omega}\times\boldsymbol{v}'(t)=$$
$$=-2\omega\left(\frac{\cos\big(\varphi(t)\big)\rho'(t)}{\rho(t)}\,\boldsymbol{v}_1(t)+\cos\big(\varphi(t)\big)\theta'(t)\,\boldsymbol{\psi}(t)+\varphi'(t)\,\sin\big(\varphi(t)\big)\,\boldsymbol{\chi}(t)\right).$$

If a particle at $\boldsymbol{v}(t)$ is moving radially away from the center of the earth, the Coriolis force on it is $-2\omega\cos\big(\varphi(t)\big)\rho'(t)/\rho(t)$ times the vector $\boldsymbol{v}_1$, which points eastward in a plane perpendicular to the north-south axis. Thus it tends to deflect particles westward as they move upward ($\rho'(t)>0$) and eastward as they move downward ($\rho'(t)<0$). In terms of a plane that is perpendicular to $\boldsymbol{v}(t)$ and for a falling particle, this vector points southeast in the northern hemisphere and northeast in the southern hemisphere.

For particles moving horizontally in terrestrial coordinates, that is, perpendicular to $\boldsymbol{v}(t)$ ($\rho'(t)=0$, so that $\rho(t)=\rho_0=$ const.), we consider separately those that are moving eastward ($\varphi'(t)=0$, so that $\varphi(t)=\varphi_0=$ const.) and those that are moving northward ($\theta'(t)=0$, so that $\theta(t)=\theta_0=$ const.).

In the case of eastward motion, the Coriolis force is $-2\omega\cos(\varphi_0)\theta'(t)\boldsymbol{\psi}(t)$, which tends to deflect the particle *directly away* from the north-south axis. The projection of this force into a plane tangent to the sphere about the origin through $\boldsymbol{v}(t)$ thus points toward the equator, so that the horizontal motion it induces is southward in the northern hemisphere and northward in the southern hemisphere. In any case, *the component of the deflection tangent to the sphere is toward the right as seen by someone looking along the direction of the tangential component of the velocity*[9] *when the particle is in the northern hemisphere, and to the left in the southern hemisphere.* There is no tangential deflection for particles moving along the equator. The Coriolis force is a radial force in that case with no tangential component, centrifugal if the point is moving eastward and centripetal if it is moving westward.

In case of northward motion, the Coriolis force is $-2\omega\sin\big(\varphi(t)\big)\varphi'(t)\boldsymbol{\chi}(t)$, which points eastward tangentially to the sphere through $\boldsymbol{v}(t)$ if the particle is moving northward ($\varphi'(t)>0$) in the northern hemisphere ($\varphi(t)>0$) or southward in the southern hemisphere, and westward in the opposite two cases. In either case, a person looking along the tangential component of the velocity[10] would again see a deflection to the right in the northern hemisphere and to the left in the southern hemisphere. The effect is more pronounced away from the equator.

Finally, to account for the last of the three "fictitious" forces, we note that it is $-\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\boldsymbol{v})$, which, according to the work done above, is $-\omega^2\cos(\varphi)\boldsymbol{\psi}(t)$. Since $\boldsymbol{\psi}$ always points toward the north-south axis, and its magnitude is $\rho(t)$, we see that this force, when attached at $\boldsymbol{v}(t)$, points away from the north-south axis around which the particle is revolving, as seen in celestial coordinates, and its magnitude is $\big(\rho(t)\cos\big(\varphi(t)\big)\big)\omega^2=v^2(t)/r(t)=v(t)r(t)$, where $v=\omega(t)r(t)$ is the linear speed of the particle and $r(t)=\rho(t)\cos\big(\varphi(t)\big)$ is the radius of the circle. This is precisely the acceleration of a particle in motion on a circle of radius $r(t)$ at speed $v(t)$.

[9] Assuming the person is standing on his feet at the surface of the earth and not on his head!
[10] Standing upright. See the previous note.

**4.6. Angular momentum and torque.** Let $\boldsymbol{L}_i(t)$ and $\boldsymbol{L}_r(t)$ denote the the angular momentum of a particle of mass $m$ located at the point with coordinates $\boldsymbol{u}(t)$ or $\boldsymbol{v}(t)$ about the fixed point $\mathbf{0}$, as measured in the inertial and rotating coordinate systems used respectively by Observer 1 and Observer 2. Then, suppressing the argument $t$, we have

$$\begin{aligned}
\boldsymbol{L}_i &= m\boldsymbol{u}\times\boldsymbol{u}' \\
&= mR(\boldsymbol{v})\times\big(R(\boldsymbol{v}')+R(\boldsymbol{\omega})\times R(\boldsymbol{v})\big) \\
&= R(\boldsymbol{L}_r)+m\boldsymbol{u}\times(\boldsymbol{w}\times\boldsymbol{u}) \\
&= R(\boldsymbol{L}_r)+m|\boldsymbol{u}|^2\boldsymbol{w}-m(\boldsymbol{u}\cdot\boldsymbol{w})\boldsymbol{u}\,; \\
\boldsymbol{L}_r &= R^\intercal(\boldsymbol{L}_i)+m(\boldsymbol{\omega}\cdot\boldsymbol{v})\boldsymbol{v}-m|\boldsymbol{v}|^2\boldsymbol{\omega}\,.
\end{aligned}$$

The torque $\boldsymbol{N}$ about the fixed point acting on this particle is the time derivative of the angular momentum:

$$\begin{aligned}
\boldsymbol{N}_i &= \boldsymbol{L}_i' = m\boldsymbol{u}\times\boldsymbol{u}'' \\
&= mR(\boldsymbol{v})\times\big(R(\boldsymbol{v}'')+\boldsymbol{w}'\times\boldsymbol{u}+2\boldsymbol{w}\times\boldsymbol{u}'-\boldsymbol{w}\times(\boldsymbol{w}\times\boldsymbol{u})\big) \\
&= R(\boldsymbol{N}_r)+m\boldsymbol{u}\times(\boldsymbol{w}'\times\boldsymbol{u})+2m\boldsymbol{u}\times(\boldsymbol{w}\times\boldsymbol{u}')-m\boldsymbol{u}\times\big(\boldsymbol{w}\times(\boldsymbol{w}\times\boldsymbol{u})\big) \\
&= R(\boldsymbol{N}_r)+m\boldsymbol{u}\times(\boldsymbol{w}'\times\boldsymbol{u})+2m\boldsymbol{u}\times(\boldsymbol{w}\times\boldsymbol{u}')+m(\boldsymbol{u}\cdot\boldsymbol{w})(\boldsymbol{w}\times\boldsymbol{u})\,.
\end{aligned}$$

We also have

$$\begin{aligned}
\boldsymbol{N}_i &= R(vN_r)+m\boldsymbol{u}\times(\boldsymbol{w}'\times\boldsymbol{u})+2m\boldsymbol{u}\times\big(\boldsymbol{w}\times R(\boldsymbol{v}')\big)+m\boldsymbol{u}\times\big(\boldsymbol{w}\times(\boldsymbol{w}\times\boldsymbol{u})\big)\,, \\
\boldsymbol{N}_r &= R^\intercal(\boldsymbol{N}_i)-m\boldsymbol{v}\times(\boldsymbol{\omega}'\times\boldsymbol{v})-2m\boldsymbol{v}\times(\boldsymbol{\omega}\times\boldsymbol{v}')-m(\boldsymbol{\omega}\cdot\boldsymbol{v})(\boldsymbol{\omega}\times\boldsymbol{v})\,.
\end{aligned}$$

Observer 2 can use the Newtonian law $\boldsymbol{N}_r=\boldsymbol{L}_r'$, provided the two sides of this equation are defined by the equalities just derived.

Notice in particular that if one of the observers finds no torque acting on the particle, the other *will*, in general, find some torque. If, for example, the particle is a molecule of a rotating rigid body in which the coordinate axes of Observer 2 are fixed, and therefore motionless as far as Observer 2 is concerned ($\boldsymbol{v}'=\mathbf{0}$), there will nevertheless be a torque on that particle equal to

$$\begin{aligned}
\boldsymbol{N}_i &= m\boldsymbol{u}\times(\boldsymbol{w}'\times\boldsymbol{u})+2m\boldsymbol{u}\times(\boldsymbol{w}\times\boldsymbol{u}')+m(\boldsymbol{u}\cdot\boldsymbol{w})(\boldsymbol{w}\times\boldsymbol{u}) \\
&= m\boldsymbol{u}\times(\boldsymbol{w}'\times\boldsymbol{u})+2m\boldsymbol{u}\times\big(\boldsymbol{w}\times\big(R(\boldsymbol{v}')+\boldsymbol{w}\times\boldsymbol{u}\big)+m(\boldsymbol{u}\cdot\boldsymbol{w})(\boldsymbol{w}\times\boldsymbol{u}) \\
&= m\boldsymbol{u}\times(\boldsymbol{w}'\times\boldsymbol{u})+2m\boldsymbol{u}\times\big(\boldsymbol{w}\times(\boldsymbol{w}\times\boldsymbol{u})\big)+m(\boldsymbol{u}\cdot\boldsymbol{w})(\boldsymbol{w}\times\boldsymbol{u}) \\
&= m\boldsymbol{u}\times(\boldsymbol{w}'\times\boldsymbol{u})-m(\boldsymbol{u}\cdot\boldsymbol{w})(\boldsymbol{w}\times\boldsymbol{u})\,.
\end{aligned}$$

If the body is rotating as a "sleeping top," for example ($\boldsymbol{w}'=\mathbf{0}$), this torque is $-m(\boldsymbol{u}\cdot\boldsymbol{w})(\boldsymbol{w}\times\boldsymbol{u})$. The force on a particle of mass $m$ at the point $\boldsymbol{u}$, which is rotating about its projection on the instantaneous axis of rotation with angular speed $\omega$, is therefore

$$-m\omega^2\Big(\boldsymbol{u}-\frac{\boldsymbol{u}\cdot\boldsymbol{\omega}}{\omega^2}\boldsymbol{\omega}\Big)\,.$$

This is precisely the centripetal force for such a particle in circular motion at that distance from the axis of rotation. That force is directly measurable only in the inertial system. In the rotating system, this centripetal force must be regarded as canceling the centrifugal force needed in any rotating coordinate system. From the point of view of the inertial observers, the centripetal force that keeps the particle from flying off is due to the the internal cohesive forces in the rigid body.

**4.7. Rigid bodies.** We now consider a physical system consisting of $p$ particles of masses $m_1, \dots, m_p$ and located at points with inertial coordinates $\boldsymbol{u}_1(t), \dots, \boldsymbol{u}_p(t)$ at time $t$. For the time being, we shall use only inertial coordinates.

If the particle at $\boldsymbol{u}_k(t)$ has mass $m_k$, the total mass of the body is $M = m_1 + \cdots + m_p$. The center of mass of the body $\boldsymbol{u}_c(t)$ is given by

$$\boldsymbol{u}_c(t) = \frac{1}{M} \sum_{k=1}^{p} m_k \boldsymbol{u}_k(t)\,,$$

from which it follows that

$$\begin{aligned} M\boldsymbol{u}_c'(t) &= \sum_{k=1}^{p} m_k \boldsymbol{u}_k'(t)\,, \\ M\boldsymbol{u}_c''(t) &= \sum_{k=1}^{p} m_k \boldsymbol{u}_k''(t)\,. \end{aligned}$$

That is to say, *the center of mass of the body moves exactly as a particle of mass $M$ equal to the sum of the masses of the particles of the body and concentrated at that point would move if subjected to the force that is the vector sum of all the forces acting on the individual particles of the body, and the momentum of such a particle would be the vector sum of the momenta of the individual masses that make up the body.*

Let us now consider the effect of applying a force $\boldsymbol{F}$ to a point $\boldsymbol{u}$ of the rigid body. Obviously, it produces a torque $\boldsymbol{u} \times \boldsymbol{F}$ about the origin of the inertial coordinate system. We ask what the resultant effect on the rigid body will be if we apply two such forces at two different points? Since the component of the force $\boldsymbol{F}$ parallel to $\boldsymbol{u}$ produces translation relative to the origin, it is obvious that the resultant of those components is equivalent to applying their vector sum at the origin. That leaves the components perpendicular to the points $\boldsymbol{u}$ to be considered. These components tend to produce rotation, and the magnitude and direction of this tendency is measured by the torque $\boldsymbol{N} = \boldsymbol{u} \times \boldsymbol{F}$. To find the resultant of two torques, we need to go all the way back to Archimedes' law of the lever and then apply a bit of geometry.

Suppose then that we apply two forces $\boldsymbol{F}$ and $\boldsymbol{G}$ simultaneously at points $\boldsymbol{u}$ and $\boldsymbol{v}$. We can assume that the directions of $\boldsymbol{F}$ and $\boldsymbol{G}$ are perpendicular to $\boldsymbol{u}$ and $\boldsymbol{v}$ respectively, since only the component of the force tangential to the sphere through $\boldsymbol{u}$ about the origin will produce any torque about the origin. (The normal component produces translation.) The torques are represented as the vectors $\boldsymbol{u} \times \boldsymbol{F}$ and $\boldsymbol{v} \times \boldsymbol{G}$, and their magnitudes are $|\boldsymbol{u}|\,|\boldsymbol{F}|$ and $|\boldsymbol{v}|\,|\boldsymbol{G}$. By Archimedes' law of the lever (or by incontestable algebra in our case, although one needs to "dedimensionalize" $|\boldsymbol{u}|$ to avoid absurdity),[11] we have

$$\frac{1}{|\boldsymbol{u}|}\boldsymbol{u} \times (|\boldsymbol{u}|\,\boldsymbol{F}) = \boldsymbol{u} \times \boldsymbol{F}\,, \quad \frac{1}{|\boldsymbol{v}|}\boldsymbol{v} \times (|\boldsymbol{v}|\,\boldsymbol{G}) = \boldsymbol{v} \times \boldsymbol{G}\,.$$

Hence, if we replace $\boldsymbol{F}$ by $\boldsymbol{F}_1 = |\boldsymbol{u}|\,\boldsymbol{F}$ and $\boldsymbol{G}$ by $\boldsymbol{G}_1 = |\boldsymbol{v}|\boldsymbol{G}$, applying the first of these at $\boldsymbol{u}_1 = \boldsymbol{u}/|\boldsymbol{u}|$ and the second at $\boldsymbol{v}_1 = \boldsymbol{v}/|\boldsymbol{v}|$, both of which are at unit

[11] We are invoking the privilege of using the same coordinate system for objects of different physical dimension. As long as we keep units consistent in line with the discussion at the beginning of this section, and the spatial coordinates underlying the system do not change, that will lead to no contradictions.

distance from the origin, we will get the same two torques about the origin. Finally, the force $\boldsymbol{F}_1$ applied at $\boldsymbol{u}_1$ can be replaced by a force $\boldsymbol{F}_2$ of equal magnitude applied at $\boldsymbol{u}_2$, where $\boldsymbol{u}_2$ is any other point of the unit circle through $\boldsymbol{u}_1$ in the plane of $\boldsymbol{u}_1$ and $\boldsymbol{F}_1$, provided $\boldsymbol{F}_2$ is also tangent to this circle and gives it the same orientation. (In the language used above, $\boldsymbol{F}_2$ is obtained by parallel transport of $\boldsymbol{F}_1$ along this circle.) And similarly, $\boldsymbol{G}_1$ applied at $\boldsymbol{v}_1$ can be replaced by $\boldsymbol{G}_2$ applied at $\boldsymbol{v}_2$. Thus, we finally have

$$\boldsymbol{u} \times \boldsymbol{F} = \boldsymbol{u}_2 \times \boldsymbol{F}_2\,, \quad \boldsymbol{v} \times \boldsymbol{G} = \boldsymbol{v}_2 \times \boldsymbol{G}_2\,.$$

Since the two unit circles through $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ intersect, we can choose $\boldsymbol{u}_2 = \boldsymbol{v}_2$ to be one of the two points of intersection. The resultant of the two torques is now clear. It is

$$\boldsymbol{u}_2 \times (\boldsymbol{F}_2 + \boldsymbol{G}_2) = \boldsymbol{u}_2 \times \boldsymbol{F}_2 + \boldsymbol{v}_2 \times \boldsymbol{G}_2 = \boldsymbol{u} \times \boldsymbol{F} + \boldsymbol{v} \times \boldsymbol{G}\,.$$

In short, the resultant of two torques is simply their vector sum, showing once again the elegance with which vector analysis is in harmony with Newtonian mechanics.

**4.8. A rigid body with a fixed point.** Now restoring the assumption that the two coordinate systems share a common origin at a point of the rigid body that is fixed in space, we see that the total torque acting on the body, as judged by Observer 1, is

$$\boldsymbol{N}_i(t) = \sum_{k=1}^{p} m_k\big(u_k(t) \times u_k''(t)\big) = \sum_{k=1}^{p} \boldsymbol{L}_{ki}'(t)\,.$$

If we define the angular momentum $\boldsymbol{L}_i(t)$ of the body about the fixed point to be the sum of the angular momenta of the individual particles about that point, we have once again

$$\boldsymbol{N}_i(t) = \boldsymbol{L}_i'(t)\,.$$

The crucial fact we now invoke is that, *in the rotating coordinate system, every point of the body is at rest.* In terms of the inertial system, this means that the total angular momentum is

$$\begin{aligned}
\boldsymbol{L}_i(t) &= \sum_{k=1}^{p} m_k \boldsymbol{u}_k(t) \times \boldsymbol{u}_k'(t) \\
&= \sum_{k=1}^{p} m_k \boldsymbol{u}_k(t) \times \big(R(t)\boldsymbol{v}_k' + \boldsymbol{w}(t) \times \boldsymbol{u}_k(t)\big) \\
&= \sum_{k=1}^{p} m_k \boldsymbol{u}_k(t) \times \big(\boldsymbol{w}(t) \times \boldsymbol{u}_k(t)\big)\,.
\end{aligned}$$

It follows that

$$\begin{aligned}
R^{\tau}(t)\boldsymbol{L}_i(t) &= \sum_{k=1}^{p} m_k \boldsymbol{v}_k \times \big(\boldsymbol{\omega}(t) \times \boldsymbol{v}_k\big) \\
&= \sum_{k=1}^{p} m_k |\boldsymbol{v}_k|^2 \boldsymbol{\omega}(t) - \sum_{k=1}^{p} m_k (\boldsymbol{\omega}(t) \cdot \boldsymbol{v}_k)\boldsymbol{v}_k\,.
\end{aligned}$$

Consideration of the individual components of this vector, assuming $\boldsymbol{v}_k = (x^k, y^k, z^k)$, leads to the conclusion that

$$R^\tau(t)\boldsymbol{L}_i(t) = \mathfrak{I}\big(\boldsymbol{\omega}(t)\big)\,,$$

where $\mathfrak{I}$ is the linear operator whose matrix in the rotating coordinate system is the symmetric matrix

$$\begin{pmatrix} \sum_k m_k\big((y^k)^2+(z^k)^2\big) & -\sum_k m_k x^k y^k & -\sum_k m_k x^k z^k \\ -\sum_k m_k x^k y^k & \sum_k m_k\big((x^k)^2+(z^k)^2\big) & -\sum_k m_k y^k z^k \\ -\sum_k m_k x^k z^k & -\sum_k m_k y^k z^k & \sum_k m_k\big((x^k)^2+(y^k)^2\big) \end{pmatrix}.$$

Since the operator $\mathfrak{I}$ has a symmetric matrix in any given orthonormal basis, it is diagonalizable. That is, we can choose the body axes so that the "products of inertia" such as $-\sum_k m_k x^k z^k$ are all zero. This diagonalizability theorem was proved by Euler.[12] Euler's result is basically the spectral theorem, which asserts that there exists an orthonormal basis of $\mathbb{R}^3$ in which the matrix of the operator represented by such a matrix is diagonal.[13]

Moreover, as one can easily compute, if $\boldsymbol{\omega} \neq \boldsymbol{0}$, then

$$\boldsymbol{\omega}\cdot\mathfrak{I}(\boldsymbol{\omega}) = \sum_{k=1}^{p} m_k\big(|\boldsymbol{v}_k|^2|\boldsymbol{\omega}|^2 - (\boldsymbol{\omega}\cdot\boldsymbol{v}_k)^2\big)\,,$$

which, by the Schwarz inequality, is positive except for a few very trivial cases in which each term in this sum can be zero for some non-zero value of $\boldsymbol{w}$.

Thus, we can assume that the axes of the rotating coordinate system have been chosen so that the matrix of the operator $\mathfrak{I}$ is diagonal and has positive numbers, say $A$, $B$, and $C$ on its main diagonal.

To generalize what we have done to the case of a body having a continuous density $\rho$, simply replace the entries in its matrix by

$$\iiint_B (x^2+y^2)\rho(x,y,z)\,dx\,dy\,dz\,, \quad -\iiint_B xy\rho(x,y,z)\,dx\,dy\,dz\,,$$

and so on.

The matrix of the operator $\mathfrak{I}$ is constant in the rotating coordinates that are fixed in the body.

We conclude that

$$\boldsymbol{L}_i(t) = R(t)\mathfrak{I}R^\tau(t)\big(\boldsymbol{w}(t)\big) = \mathfrak{J}(t)\big(\boldsymbol{w}(t)\big)\,,$$

where $\mathfrak{J}(t) = R(t)\mathfrak{I}R^\tau(t)$.

Then, from simple algebra, the torque on the body is

$$\boldsymbol{N}_i(t) = \boldsymbol{L}_i'(t) = R'(t)\mathfrak{I}R^\tau(t)\big(\boldsymbol{w}(t)\big) + R(t)\mathfrak{I}(R^\tau)'(t)\big(\boldsymbol{w}(t)\big) + R(t)\mathfrak{I}R^\tau(t)\big(\boldsymbol{w}'(t)\big)\,.$$

---

[12] See his books *Mechanica, sive motus scientia analytice exposita* (*Mechanics Expounded Analytically as the Science of Motion*, Petersburg Academy of Sciences, 1736) and *Introductio in analysin infinitorum* (*Introduction to Infinitesimal Analysis*, M. M. Bousquet, Lausanne, 1758). He says that he was led to this discovery by pondering the question whether a rigid body subject to no forces could rotate about a constant axis; it turns out that this is the case, but only for the three principal axes of the body. A homogeneous ball, of course, has more than just three such axes; any diameter is such an axis.

[13] For the author's favorite elementary proof of this fact, see the paper "An uncharacteristic proof of the spectral theorem," written by the same author, and published in *Mathematics Magazine*, **52** (1979), No. 5, 302–304.

Replacing $\boldsymbol{w}(t)$ by $R(t)\big(\boldsymbol{\omega}(t)\big)$ and $\boldsymbol{w}'(t)$ by $R(t)\big(\boldsymbol{\omega}'(t)\big)$, then multiplying by $R^\tau(t)$ we get

$$R^\tau(t)\big(\boldsymbol{N}_i(t)\big) = \boldsymbol{\omega}(t) \times \mathfrak{I}\big(\boldsymbol{\omega}(t)\big) + \mathfrak{I}\big(-\boldsymbol{\omega}(t) \times \boldsymbol{\omega}(t)\big) + \mathfrak{I}\big(\boldsymbol{\omega}'(t)\big)\,.$$

As the middle term on the right is zero, we finally get an equation for the angular velocity $\boldsymbol{\omega}'(t)$:

$$\boldsymbol{\Lambda}(t) = R^\tau(t)\big(\boldsymbol{N}_i(t)\big) = \boldsymbol{\omega}(t) \times \mathfrak{I}\big(\boldsymbol{\omega}(t)\big) + \mathfrak{I}\big(\boldsymbol{\omega}'(t)\big)\,.$$

As a way of finding $\boldsymbol{\omega}(t)$, this equation seems at first sight to be a case of circular reasoning, since it contains $R^\tau(t)$ as data. Obviously, if we knew $R^\tau(t)$, we would *already* be able to compute $\boldsymbol{\omega}(t) = R^\tau(t)R'(t)$. But there are a number of special cases in which the equation becomes solvable. In the most important of these cases, the torque $\boldsymbol{N}_i(t)$ is provided by gravity. That is, the torque acting on the particle of mass $m_n$, at $\boldsymbol{u}_n(t)$ is $\boldsymbol{N}_n(t) = \boldsymbol{u}_n(t) \times (-m_n g\boldsymbol{k})$, where $g$ is the acceleration of gravity. The total torque is

$$\boldsymbol{N}_i(t) = g\boldsymbol{k} \times \sum_{n=1}^{p} m_n \boldsymbol{u}_n(t) = g\boldsymbol{k} \times M\boldsymbol{u}_c(t)\,,$$

where $\boldsymbol{u}_c(t)$ is the location of the center of mass of the body and $M$ is the total mass of the body.

We now have

$$\boldsymbol{\Lambda}(t) = R^\tau(t)\big(\boldsymbol{N}_i(t)\big) = -Mg\boldsymbol{v}_c \times \boldsymbol{\gamma}(t)\,,$$

where $\boldsymbol{\gamma}(t)$ is the unit vector $R^\tau(t)(\boldsymbol{k})$.

In other words, the components of $\boldsymbol{\gamma}(t)$ are just $r_{31}(t)$, $r_{32}(t)$, $r_{33}(t)$, so that we don't actually have to know *all* of $R^\tau(t)$, only its last column. Notice that $\boldsymbol{\gamma}'(t) = (R^\tau)'(t)(\boldsymbol{k})$, so that $R(t)\big(\boldsymbol{\gamma}'(t)\big) = \big(R(t)(R^\tau)'(t)\big)(\boldsymbol{k}) = \boldsymbol{k} \times \boldsymbol{w}$ and we get the system of differential equations

$$\begin{aligned} \mathfrak{I}\big(\boldsymbol{\omega}'(t)\big) &= \mathfrak{I}\big(\boldsymbol{\omega}(t)\big) \times \boldsymbol{\omega}(t) - Mg\boldsymbol{v}_c \times \boldsymbol{\gamma}(t)\,, \\ \boldsymbol{\gamma}'(t) &= \boldsymbol{\gamma}(t) \times \boldsymbol{\omega}\,. \end{aligned}$$

The theoretical solution to the mechanical problem is now complete. What remains is the purely mathematical problem of solving this system of six differential equations. That mathematical problem in general is formidable.

**4.9. The spinning top.** Let us denote the components of $\boldsymbol{\omega}(t)$ by $p(t)$, $q(t)$, and $r(t)$ and those of $\boldsymbol{v}_c$ by $v_{c1}$, $v_{c2}$, and $v_{c3}$. We have already noted that the components of $\boldsymbol{\gamma}(t)$ are $r_{31}(t)$, $r_{32}(t)$, and $r_{33}(t)$. We get the following system of six differential equations, first written down in this form by the British mathematician

Robert Baldwin Hayward (1829–1903) in 1858:[14]

$$\begin{aligned}
A\frac{dp}{dt} &= (B-C)qr - Mg(\upsilon_{c2}r_{33} - \upsilon_{c3}r_{32})\,,\\
B\frac{dq}{dt} &= (C-A)rp - Mg(\upsilon_{c3}r_{31} - \upsilon_{c1}r_{33})\,,\\
C\frac{dr}{dt} &= (A-B)pq - Mg(\upsilon_{c1}r_{32} - \upsilon_{c2}r_{31})\,,\\
\frac{dr_{31}}{dt} &= r_{32}r - r_{33}q\,,\\
\frac{dr_{32}}{dt} &= r_{33}p - r_{31}r\,,\\
\frac{dr_{33}}{dt} &= r_{31}q - r_{32}p\,.
\end{aligned}$$

This being an autonomous system, in which the right-hand sides do not depend explicitly on $t$, we can eliminate $dt$ and get a system of five differential equations, all the following differentials being equal to $dt$:

$$\frac{A\,dp}{P} = \frac{B\,dq}{Q} = \frac{C\,dr}{R} = \frac{dr_{31}}{R_{31}} = \frac{dr_{32}}{R_{32}} = \frac{dr_{33}}{R_{33}}\,.$$

Here, of course, $P$, $Q$, $R$, $R_{31}$, $R_{32}$, and $R_{33}$ are the right-hand sides of the equations above. To solve this system, we need to find five independent integrals, that is, five conservation laws. But, because of the obvious equations

$$0 = \frac{\partial P}{\partial p} = \frac{\partial Q}{\partial q} = \frac{\partial R}{\partial r} = \frac{\partial R_{31}}{\partial r_{31}} = \frac{\partial R_{32}}{\partial r_{32}} = \frac{\partial R_{33}}{\partial r_{33}}\,,$$

the divergence condition of Jacobi's last-multiplier principle is satisfied, that is,

$$\frac{\partial P}{\partial p} + \frac{\partial Q}{\partial q} + \frac{\partial R}{\partial r} + \frac{\partial R_{31}}{\partial r_{31}} + \frac{\partial R_{32}}{\partial r_{32}} + \frac{\partial R_{33}}{\partial r_{33}} = 0\,.$$

Because this condition holds, we need to aim at getting four independent integrals. Jacobi's method will then do the rest. We start out hopefully.

1. Obviously, since $\boldsymbol{\gamma}(t)$ is a unit vector, we have $f_1(p,q,r,r_{31},r_{32},r_{33}) = r_{31}^2 + r_{32}^2 + r_{33}^2 = 1$.
2. Conservation of energy tells us that the sum of the kinetic and potential energies is constant. Thus

$$f_2(p,q,r,r_{31},r_{32},r_{33}) = \frac{1}{2}(Ap^2 + Bq^2 + Cr^2) + Mg(\upsilon_{c1}r_{31} + \upsilon_{c2}r_{32} + \upsilon_{c3}r_{33}) = E\,.$$

3. Because the force is vertical, the torque $\boldsymbol{N}_1$ is horizontal, and therefore the vertical component of the angular momentum is constant, that is

$$f_3(p,q,r,r_{31},r_{32},r_{33}) = Apr_{31} + Bqr_{32} + Crr_{33} = \Omega\,.$$

Although we have used geometrical and physical principles to state these conservation laws, they are immediately computable. That is, the equations imply easily that $df_1 = df_2 = df_3 = 0$. For example, $df_1 = 2\boldsymbol{\gamma}\cdot\boldsymbol{\gamma}'\,dt = 2\boldsymbol{\gamma}\cdot(\boldsymbol{\gamma}\times\boldsymbol{\omega})\,dt = 0$. The conservation of the other two quantities can also be computed directly (Problem 5.10).

[14] "On a direct method of estimating velocities, accelerations, and all similar quantities with respect to axes movable in any manner in space," *Transactions of the Cambridge Philosophical Society*, **X**: 1–22.

## 5. The Exactly Solvable Cases

Since we are unable to find a fourth integral, we can't expect to solve the general system. Because of its symmetry and the fact that three integrals lie near the surface, this problem is as attractive as the proverbial Lorelei on the river Rhine, the siren that in legend lured sailors to their doom. Indeed, Sof'ya Kovalevskaya (1850–1891), who found the third and last of the exactly solvable cases of this motion, said that the German mathematicians called this problem the "mathematical mermaid" (die mathematische Nixe). As a mathematical goal, it can be equally deadly. The Prussian Academy of Sciences offered a prize for its solution during the 1850's, but there were no entries to the competition, and it turns out that there are only three special cases in which the angular velocity is a meromorphic function of time (a function having only poles as singularities when time is regarded as a complex variable).

**5.1. The Lagrange case.** In the case of actually useful spinning tops, such as the gyroscopes once used in navigation systems, the problem has some extra symmetry that makes it possible to find a fourth integral. This case was studied by Joseph-Louis Lagrange (Giuseppe-Lodovico Lagrangia, 1736–1813), and is found in Section IX, § 11, of the 1811 edition of his classic work *Mécanique analytique*. There, it is assumed that the body has cylindrical symmetry, so that $A = B$, and that the fixed point lies on the axis of symmetry, which also contains the center of mass, so that $\boldsymbol{\upsilon}_{c1} = 0 = \boldsymbol{\upsilon}_{c2}$. In that case, the third equation becomes very simple, and we get a leg up on the problem by invoking the integral

$$f_0(p, q, r, r_{31}, r_{32}, r_{33}) = r\,,$$

which is constant in this case. Moreover, by simply freezing $r$ at a constant value and removing the equation involving $dr$ from the system we find that the integrating factor needed to retain the Jacobi criterion is simply 1. That is, the remaining system of four equations in five unknowns still meets the divergence criterion. We can then go to work on it by successively eliminating variables using $f_1$, $f_2$, and $f_3$ until we are finally left with an exact equation in two variables. Since the integrals are all polynomial functions, the Jacobi method guarantees that the solution can be expressed using integrals of algebraic functions. That is, the solution is what is called an *abelian* function of the time variable. In the Lagrange case, the abelian functions are elliptic functions.

**5.2. The Euler case.** If we restore the generality of the body by eliminating the assumption that $A = B$, but place the fixed point at the center of mass ($\boldsymbol{\upsilon}_c = \mathbf{0}$), we get what is known as the Euler case of the motion. The original six equations of motion then separate into two sets of three equations each. When when $dt$ is eliminated from the first set of three, the result is a pair of equations that is divergence-free. Conservation of energy then provides one integral, which is all that is needed to invoke the last-multiplier principle. Once we have $p$, $q$, and $r$, the fact that the $\boldsymbol{\gamma}$ is a unit vector points gives us the only integral we need to find $r_{31}$, $r_{32}$, and $r_{33}$. As in the Lagrange case, the general solution can be expressed in terms of elliptic functions.

**5.3. The Kovalevskaya case.** In both the Euler and Lagrange cases we impose three extra conditions on the data to get a solution by quadratures. In the only other case in which solution by quadratures is possible, the conditions $A = B = 2C$,

$\upsilon_3 = 0$ are imposed. This case, in which the motion is very erratic, was discovered by Sof'ya Kovalevskaya (1850–1891) in 1886, and in 1888, she won a prize from the Paris Academy of Sciences for the achievement. One can assume that the center of mass lies on one of the axes in the plane perpendicular to the third body axis, since the axes in this plane can be rotated without changing the equations. Then, by a suitable choice of units, one can assume that $A = B = 2$ and $C = 1$. Under those assumptions, the fourth integral was given by Kovalevskaya as

$$f(p, q, r, r_{31}, r_{32}, r_{33}) = (p^2 - q^2 - Mg\upsilon_{c1}r_{31})^2 + (Mg\upsilon_{c1}r_{32} - 2pq)^2 .$$

(She actually gave it in a simpler form by using complex numbers.) In this third case also, the integrals are all polynomial functions of the variables, and so the solution can be expressed using integrals of algebraic functions. In this case, however, the solutions were *hyperelliptic* integrals, which involved the square root of a polynomial of degree 5. To express the solution explicitly, one needs theta functions of two variables, which were created during the 1840s and 1850s, and their properties developed by Riemann and Weierstrass. Kovalevskaya's paper was a new physical application of these recently-discovered functions, and that was the reason the Paris Academy of Sciences rewarded her with the 1888 Bordin Prize. Her remark that hers was the last case in which the general solution was a meromorphic function of time was not completely proved in her paper; the full proof was given by Aleksandr Mikhailovich Lyapunov (1857–1918) in 1894.

## 6. Problems

**Problem 5.1.** Prove the basic facts about integrals of vector-valued functions, that is, that the norm of the integral is not greater than the integral of the norm and that the derivative of an indefinite integral is the integrand.

**Problem 5.2.** Consider the case $p = 1$ (real-valued functions of a real variable) and the differential equation $y'(x) = y(x)$ with initial condition $y(0) = 1$. Execute the method of successive approximations on this initial-value problem and express the solution of it as convergent infinite series. What function does this series represent?

**Problem 5.3.** Consider an $n$th-order initial value problem for a real-valued function $y(x)$ consisting of a differential equation

$$y^{(n)}(x) = f\big(x, y(x), y'(x), y''(x), \ldots, y^{(n-1)}(x)\big)$$

with initial conditions $y(x_0) = y^0$, $y'(x_0) = y^1$,..., $y^{(n-1)}(x_0) = y^{n-1}$, where $f$ is a real-valued function having continuous partial derivatives on a connected open subset $U$ of $\mathbb{R}^{n+1}$ containing the point $(x_0, y^0, y^1, \ldots, y^{n-1})$. Prove that this problem has a unique solution if the partial derivatives of $f$ on all of its variables except the first are uniformly bounded in some cube $C$ in $\mathbb{R}^{n+1}$ containing the point $(x_0, y^0, y^1, \ldots, y^{n-1})$ and contained in $U$. *Hint:* Let $\boldsymbol{y}(x) = \big(y(x), y'(x), \ldots, y^{(n-1)}(x)\big)$.

**Problem 5.4.** The space of continuous functions on an interval $[a, b]$ of the real line with values in $\mathbb{R}^p$ is a complete metric space under the norm

$$|\boldsymbol{y}|_\infty = \sup_{a \le x \le b} |\boldsymbol{y}(x)| .$$

(The proof of this fact is the simple observation that convergence of a sequence $\{\boldsymbol{y}_n\}_{n=0}^{\infty}$ in this norm means uniform convergence of each component $y^{nj}$, $1 \le j \le p$.)

Prove that if $X$ is any complete metric space and $\varphi : X \to X$ is a *contraction*, that is, there exists a positive number $\alpha < 1$ such that for all $x \in X$ and $y \in X$, $d\big(\varphi(x), \varphi(y)\big) \le \alpha d(x, y)$, then the mapping $\varphi$ has a unique fixed point $z$ such that $\varphi(z) = z$, and it can be reached starting from any point $x_0$ by setting $x_{n+1} = \varphi(x_n)$, $n = 0, 1, \ldots$ and taking the limit as $n \to \infty$. Show how to derive our existence and uniqueness theorem from this fixed-point theorem by constructing a suitable $X$ and $\varphi$.

**Problem 5.5.** Show that the solution of the initial-value problem

$$\boldsymbol{y}(x) = \boldsymbol{y}_0 + \int_{x_0}^{x} \boldsymbol{f}\big(t, \boldsymbol{y}(t)\big)\, dt$$

actually has continuous derivatives of order $k+1$ provided all the partial derivatives of $\boldsymbol{f}(x, \boldsymbol{y})$ of order $k$ or less are continuous.

**Problem 5.6.** Prove that the system of differential equations

$$\frac{\left(\frac{\partial f}{\partial x^{n+1}}\right)^* dx^1}{M_1^*} = \cdots = \frac{\left(\frac{\partial f}{\partial x^{n+1}}\right)^* dx^n}{M_n^*}$$

satisfies the vanishing-divergence condition.

**Problem 5.7.** Consider a particle of mass $M$ fixed in space and another of mass $m$ going around it in a circular orbit of radius $r$ in accordance with Newtonian physics. If the particle is at position $\boldsymbol{u}(t)$ at time $t$ in a system of coordinates fixed in absolute space with origin at the particle of mass $M$, then the force on the orbiting particle is a purely centripetal force

$$\boldsymbol{F}(t) = -\frac{GMm}{|\boldsymbol{u}(t)|^3}\boldsymbol{u}(t) = -\frac{GMm}{r^3}\boldsymbol{u}(t)\,.$$

Now imagine a second set of coordinates, again with origin at the particle of mass $M$, but rotating in such a way that the position of the particle of mass $m$ is constantly $\boldsymbol{v}$. Show that there will be no Coriolis force acting on the orbiting particle in this system, but there will be a centrifugal force equal and opposite to the centripetal force in the other system, and therefore the net force on the now-stationary particle is zero. (Use Kepler's third law, which states that the period of revolution is

$$T = \frac{2\pi|\boldsymbol{v}|^{3/2}}{\sqrt{GM}} = \frac{2\pi r^{3/2}}{\sqrt{GM}}\,,$$

to get the angular velocity vector $\boldsymbol{\omega}$.)

**Problem 5.8.** Suppose a particle moves westward along the earth's equator. What linear speed will cause the Coriolis force to cancel the centrifugal force? (The earth's angular velocity of rotation is $7.29 \times 10^{-5}$ radians per second. Take the altitude of the particle (radius of the earth) to be $6.37 \times 10^6$ meters.)

**Problem 5.9.** Carry out the computation that shows the conservation of energy and the vertical component of angular momentum in the general case of a rotating rigid body subject only to gravitational forces.

**Problem 5.10.** Use the law of conservation of energy as the integral and follow the last-multiplier principle to solve the first three equations in the Euler case in closed form.

**Problem 5.11.** Combining the Euler and Lagrange cases by assuming that $\boldsymbol{\upsilon}_c = \mathbf{0}$ and $A = B$, show that the solution in this case can be expressed in terms of elementary functions. (The rotation of the earth falls under this case, to a fair degree of approximation.)

**Problem 5.12.** To solve a second-order linear homogeneous[15] differential equation $y''(x) + F(x)y'(x) + G(x)y(x) = 0$, let $z(x) = y'(x)$ and write the equation as a system of two equations:

$$\begin{aligned} y'(x) &= z(x)\,, \\ z'(x) &= -G(x)y(x) - F(x)G(x)\,. \end{aligned}$$

Use the method of successive approximations to produce the solution of the initial-value problem $y''(x) + y(x) = 0$, $y(0) = 1$, $y'(0) = 0$.

[15] The word *homogeneous* here should not be confused with the use of that word in classifying first-order differential equations, as was done above. All it means is that the "forcing function" that would ordinarily appear on the right-hand side is zero.

**APPENDIX 6**

# Invariance

> *In practice, there must be* some *principle on which co-ordinates are assigned, and this principle must have* some *physical significance. But we might, for instance, measure time by the worst clock ever made, provided it only went wrong and did not actually stop... We want to discover a sufficient, if not necessary, condition which, if fulfilled, insures that a statement in terms of co-ordinates has a meaning independent of co-ordinates.*
>
> Bertrand Russell, *The Analysis of Matter* (1927). Reprint: Routledge, London and New York, 2001, pp. 63–64.

> *The method of tensors first assigns co-ordinates, and then shows how to obtain results which, though expressed in terms of co-ordinates, do not really depend on them. There must be a less indirect technique possible, in which we use no more apparatus than is logically necessary, and have a language which will only express such facts as are now expressed in the language of tensors, not such as depend on the choice of coordinates. I do not say that such a method, if discovered, would be preferable in practice, but I do say that it would give a better expression of the essential relations, and greatly facilitate the task of the philosopher. In the meantime, the method of tensors is technically delightful, and suffices for mathematical needs.*
>
> *Ibid.*, p. 71. (Russell wrote these words before the 1931 work of Kurt Gödel (1906–1978), which showed that there are always results in arithmetic that require "more apparatus" for their proof than can be derived from the axioms of the system. The effort to find a minimal apparatus that will make it possible to produce every result one might want is a hopeless task. But in one sense, the freeing of manifolds from any particular embedding in Euclidean space and the intrinsic definition of tangent vector fields on them might be regarded as the fulfillment of Russell's desire.)

In the present appendix, we study a concept that is of both mathematical and metaphysical importance. Invariance is concerned with the general problem of deciding when two observers are talking about the same physical quantity, and it is an essential property for concepts to have if any meaningful discussion of them is to take place. To illustrate with a rather fanciful analogy, consider how a scholarly meeting might be arranged for a given time and place. One way to assure that the participants do not get lost trying to find the venue is to put them all in a bus and transport them. If each is arriving separately, by train, airplane, or automobile, it is necessary to provide detailed directions for each of these means of transportation; and in fact, that is how meetings are announced. When the "meeting venue" is a mathematical object, we have a similar choice. We can either prescribe a "canonical" description of the object, or allow each person a private description,

but establish common rules for reconciling those descriptions. For example, the space $\mathbb{R}^3$ *is*, by definition, the set of all ordered triples of real numbers, together with certain metric and algebraic properties that do not concern us right now. Associated with it, we have a variety of individual bases that may be used by different people for various purposes. Now, we could (and do) declare a canonical basis for this space, namely the standard basis $\boldsymbol{\varepsilon}_1 = (1,0,0)$, $\boldsymbol{\varepsilon}_2 = (0,1,0)$, $\boldsymbol{\varepsilon}_3 = (0,0,1)$. As long as everyone uses that basis, no problem arises in communication. The coordinates of a point $\boldsymbol{u} = (u^1, u^2, u^3) \in \mathbb{R}^3$ with respect to this basis are identical with its components: $\boldsymbol{u} = u^1\varepsilon_1 + u^2\varepsilon_2 + u^3\varepsilon_3$. Often, however, we have occasion to use other bases—say the basis of eigenvectors of a linear operator—and when we do, the *coordinates* of a point and its *components* are two different things. As we know, two people using different bases must establish the $3 \times 3$ transition matrix that changes coordinates from one basis to another.

Just as the railroad predated the private automobile, the mathematical analogue of common transport—Euclidean space and its algebraization by Descartes, Fermat, and Euler—predated the extensive use of the "private" parametric representations of surfaces.[1] It was with Gauss's work on differential geometry in the mid-1820s that parameters truly came into their own. They are extremely flexible, and Gauss achieved a remarkable amount of insight into the geometry of surfaces by using general coordinates. Still, the "common carrier" of Euclidean space remained as the context in which the parameters were used. It was not until the mid-twentieth century that differential geometry became essentially the intrinsic study of manifolds. The author can recall reading somewhere around 1960 that coordinate systems (parameters) were "necessary evils" in differential geometry. But when the late Richard Crittenden (1930–1996), in whose course the present author sat in 1961, heard that remark, he replied, "Well, I don't think so." It was just at that time that the abstraction of tangent vectors as derivations on $C^\infty$ appeared to have freed the study of manifolds from their embeddings in Euclidean space. Nevertheless, a manifold has to be made of *something* concrete and geometrical, even though we need not be concerned with precisely *which* Euclidean space houses our manifold. And, in fact, in the 2001 reprinting of Crittenden's elegant 1964 textbook of differential geometry, his co-author Richard Bishop remarked that, although the techniques they had emphasized were later supplanted by the Koszul[2] calculus, which removes the need for grubby computation of the Christoffel symbols, "connections on bundles have made a comeback largely due to their convenience in expressing much of mathematical physics."

In fact, we do sometimes refer to an absolute object such as a manifold. But the manifold is usually embedded in a Euclidean space of higher dimension, and discussing it in its absolute form requires us to drag along extra variables, which is an esthetic inconvenience that Russell complained of in the quotation above. We did this patiently for two-dimensional surfaces embedded in $\mathbb{R}^3$ in Chapter 5, until we finally liberated ourselves entirely from the burden of the ambient absolute space and were able to carry on the discussion on the much more efficient level of two-parameter coordinate systems. The absolute object, although it may "exist" in a mathematical sense, is not the center of attention in differential geometry. What

[1] Euler, Laplace, and others used both rectangular and polar coordinates, and in that respect anticipated the notion that an object could be described by more than one set of variables, and Euler also introduced three angular parameters to describe a rotation.

[2] Named after Jean-Louis Koszul (b. 1921).

we have is a large set of patches of a Euclidean space, each representing a set of local coordinates, and a set of transformations between pairs of them, representing changes of parameters. All that we know or need to know is contained in those changes of parameters.

Such choices of language often arise in physics. For example, the general problem of classical mechanics is to predict the future location of a given material body. This prediction involves three things: (1) system of measuring devices for mass, time, and distance; (2) theoretical concepts such as velocity, momentum, acceleration, and force obeying assumed mathematical laws; (3) the ability to perform the computations imposed by physical theory on these theoretical concepts to produce a computable formula for the location of the body at a given time.

In order for two observers to talk meaningfully with each other about where the body will be at a future time, they need to know how their clocks and measuring rods can be reconciled. In Newtonian mechanics, this problem is not complicated. Time is absolute, and they can simply synchronize their clocks. As for location, since Newtonian space is absolute (the same for both observers) and Euclidean, we may assume that each of them is using a fixed right-handed orthonormal basis to specify a point in terms of three coordinates. If they are not moving relative to each other, the coordinates $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ that they assign to a given point will be related by an affine transformation

$$\boldsymbol{u}_1 = \boldsymbol{u}_0 + R(\boldsymbol{u}_2)\,,$$

where $\boldsymbol{u}_0$ is the vector from the origin used by Observer 1 to the origin used by Observer 2, and $R$ is a rotation matrix. In fact, either observer can rotate coordinates in such a way that $R$ is the identity transformation. The constant vector $\boldsymbol{u}_0$ will drop out when the two compare velocities, since $\boldsymbol{u}_1'(t) = R\big(\boldsymbol{u}_2'(t)\big)$, and the same relation will be true for accelerations. Since they also agree about the mass $m$ of the particle, the coordinates they assign to the force, if any, acting on it will be related by $\boldsymbol{F}_1 = m\boldsymbol{u}_1''(t) = R\big(m\boldsymbol{u}_2''(t)\big) = R(\boldsymbol{F}_2)$. And again, it is possible to choose coordinates so that $R$ is simply the identity, and $\boldsymbol{F}_1 = \boldsymbol{F}_2$.

To go one step further with this procedure, suppose that the observers are in relative motion on a straight line at uniform speed, but not rotating relative to each other. Then, at time $t$, the coordinates they assign to the particle at time $t$ will be related by

$$\boldsymbol{u}_1(t) = \boldsymbol{u}_0 + t\boldsymbol{v}_1 + R\big(\boldsymbol{u}_2(t)\big)\,,$$

where $\boldsymbol{u}_0$ is now the vector from the origin used by Observer 1 to the location of the origin used by Observer 2 at time $t = 0$, and $\boldsymbol{v}_1$ is the velocity Observer 1 assigns to Observer 2. They now reconcile velocities through the equation $\boldsymbol{u}_1'(t) = \boldsymbol{v}_1 + R\big(\boldsymbol{u}_2'(t)\big)$, but accelerations and forces are reconciled just as before. Thus, two observers in mutual straight-line motion without any rotation do agree about the force and acceleration of a particle, in the sense that the coordinates they assign to it are reconciled in this way. As we saw in Chapter 3, in Newtonian mechanics, they can agree about the magnitude and direction of a force while still disagreeing as to its physical cause. Special relativity enabled them to assign the "same" electromagnetic field as the cause and treated the electric and magnetic portions symmetrically.

The more general the relative motion of the two observers, the more complicated are the formulas used to reconcile their observations. As we saw in Appendix 5, if the rotation matrix $R$ is a function of $t$, then they must introduce Coriolis and

centrifugal forces along with the accelerative force before they can agree on the mechanics of the body. These examples should be kept in mind as we proceed to the general problem of getting two observers using different parametrizations to agree that they are talking about the same thing. As with these examples, the starting point is simply location. The two observers must have a way of agreeing when two sets of coordinates $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ represent the same point in space.

The two problems we wish to consider occur in different contexts. In the first case, we imagine two observers who have laid out coordinates on a space that they are studying—it may be a surface in $\mathbb{R}^3$ or a general manifold; the problem is to see what concepts they will agree on after they perform the basic labor of computing the coordinate transformations that tell them when they are talking about the same point in space. We shall illustrate this process with the intuitive example of a surface in $\mathbb{R}^3$, and show how the first and second fundamental forms are reconciled between two observers who parameterize the surface differently. In particular, we shall show that if each of them uses the same definition of the first and second fundamental forms, even though they are using different parameters, they will agree about the length of any path on the surface, the area of any portion of the surface, and the curvature at any point of the surface. The key to doing all this is the notion of the Jacobian matrix that approximates a mapping, which leads to natural rules for identifying objects defined in terms of derivatives with respect to the parameters.

Having proved the invariance of curvature under changes of parameter, we shall pass to the more abstract setting of an $n$-dimensional manifold and show how this parameter-invariance manifests itself in the concept of tensors.

An additional benefit of this labor will be a further liberation of concepts from the Euclidean spaces in which they were originally anchored. This loosening of ties to Euclidean space occurs in several steps. To take just the two-dimensional example, we began by considering surfaces in $\mathbb{R}^3$ to define curvature. We needed this very specific space because it alone of all Euclidean spaces allows the cross product of two vectors to be defined as a third vector perpendicular to both. (Geometrically, we need a unique direction normal to each point of a plane.) We used the embedding in $\mathbb{R}^3$ to provide us with the notion of distance and surface area using the metric coefficients of first fundamental form and then to define curvature by mixing in the second fundamental form. Gauss's great achievement of defining the curvature in terms of the metric coefficients and their partial derivatives freed us from the restriction to $\mathbb{R}^3$. In fact, we could then take *any* set of metric coefficients $g_{ij}$ as our starting point. The only price to pay for beginning with the $g_{ij}$ is the need to specify how these coefficients change when a different set of parameters is used. Now in fact, curvature is determined by the Christoffel symbols alone, as we later learned. True, we computed these coefficients from the metric coefficients and their derivatives; but *any* eight coefficients can be used to produce the Riemann curvature tensor, provided they transform in the correct (non-tensor!) manner when coordinates are changed. We could actually begin with the Christoffel symbols if we wished, dispensing entirely with the metric coefficients. Again, the price of doing so is the need to specify how the Christoffel symbols transform when a different set of parameters is used. All this is more abstraction than needed for the purposes of the present book; and in all of our examples, the Christoffel symbols will be computed from the metric coefficients.

## 1. Mathematical Objects and Their Algebraic Description

Our basic model, a surface in three-dimensional space, has an objective meaning as an absolutely defined subset of absolute three-dimensional Euclidean space. For example, the sphere of radius $R$ with center at a point $P$, is described as the set of all points whose distance from $P$ is $R$. The surface and the ambient space are both absolute objects, the same for every "observer" who wishes to study them. Thus, $\mathbb{R}^3$ is the set of all triples of real numbers $(a, b, c)$, and if $P$ is the point $(0, 0, 0)$, then the (open) "northern" hemisphere of radius $R$ with center at $P$ and its "prime meridian" removed is the set of all $(a, b, c) \in \mathbb{R}^3$ satisfying the conditions

$$a^2 + b^2 + c^2 = R^2\,, \quad \max(a, |b|) > 0\,, \quad c > 0.$$

These conditions define a portion of the sphere unambiguously as a subset of $\mathbb{R}^3$ and thus provide an absolute algebraic description of that set.

Surfaces, however, are two-dimensional objects, while this description contains three variables. As Euler and Gauss (among others) showed us, it is simpler to study this object (and all other smooth surfaces as well) parametrically, describing it as the range of a continuously differentiable mapping $(u, v) \mapsto \boldsymbol{r}(u, v)$, where the parameters come from a subset of $\mathbb{R}^2$ that is particular to the person using them. The problem naturally arises of reconciling computations made using one set of parameters with those made using a different set.

**Example 6.1.** One person may parameterize the surface just described using the mapping $(u, v) \mapsto \boldsymbol{r}(u, v) = (R\cos u\,\cos v, R\cos u\,\sin v, R\sin u)$, $0 < u < \pi/2$, $-\pi < v < \pi$, while another uses $(x, y) \mapsto \boldsymbol{s}(x, y) = \big(x, y, \sqrt{R^2 - x^2 - y^2}\big)$, $x^2+y^2 < R^2$, $\max(x, |y|) > 0$.

The first thing two observers using different coordinates must do is establish a correspondence between their parameter values, that is, deciding when the pairs of parameters $(u, v)$ and $(x, y)$ correspond to the same point on the surface. In our example, we have the following change-of-variable transformations

$$\begin{aligned} u &= \arccos\Big(\frac{\sqrt{x^2+y^2}}{R}\Big)\,, \\ v &= 2\arctan\Big(\frac{y}{x+\sqrt{x^2+y^2}}\Big)\,, \\ x &= R\cos u\,\cos v\,, \\ y &= R\cos u\,\sin v\,. \end{aligned}$$

The third and fourth of these equations express $x$ and $y$ as infinitely differentiable functions of $u$ and $v$. Likewise, the inverse tangent is infinitely differentiable wherever it is defined (in this case, where $y \neq 0$), and the inverse cosine is infinitely differentiable except at $\pm 1$. Thus we need to exclude the equator, where $x^2 + y^2 = R^2$.

The absolute object (portion of a sphere) that is the apparent subject of study here plays only a minor role in the work that would proceed from this point on, and in fact can be forgotten about once we know the coordinate changes $(x, y) \leftrightarrow (\varphi, \theta)$. Each observer would define distance as the length of a geodesic, and similarly define area and curvature, using a particular set of parameters. The parameters are the main focus of attention, and the only issue is whether two observers following the same procedure, each in a given set of parameters, would get the same numbers for

distance, area, and curvature. As we shall see, if they do it using the chain rule to convert derivatives from one set of parameters to the other, they will indeed get the same numbers for these three quantities.

## 2. Reconciling Parametrizations: The Chain Rule

Although we assume that the reader knows how to use the chain rule in several variables, in order to lay out all our definitions in good order, we'll start with some elementary concepts from advanced calculus/elementary real analysis. Given that two observers reconcile their space and time coordinates by establishing a functional relationship between them: $(s; x^1, x^2, x^3) = \boldsymbol{\xi}(t; y^1, y^2, y^3)$; $(t; y^1, y^2, y^3) = \boldsymbol{\eta}(s; x^1, x^2, x^3)$, the chain rule becomes our supreme guide and commander in all infinitesimal operations, that is, those involving differentiation and integration. For every object that our two observers define or use, whether tangent vectors, or length and area, or curvature—whatever it may be—they must depend on the chain rule to tell them how to convert one language into the other. To make a poor pun, they are *constrained* by the chain rule. We shall begin, however, with two-dimensional objects, which are easier to visualize and make the general case sufficiently clear.

Everything that we are doing to deduce in this section is ultimately a consequence of the chain rule and the fact that a differentiable transformation $(u, v) \mapsto (x, y)$ from one open set in $\mathbb{R}^2$ to another is well approximated by an affine transformation: If $\boldsymbol{f}(u, v) = \big(x(u, v), y(u, v)\big)$ is differentiable, then

$$\boldsymbol{f}(u + h, v + k) = \boldsymbol{f}(u, v) + \frac{\partial \boldsymbol{f}}{\partial u} h + \frac{\partial \boldsymbol{f}}{\partial v} k + \boldsymbol{\varepsilon}(u, v; h, k)\,,$$

where $\boldsymbol{\varepsilon}(u, v; h, k)$ is defined by this equation and has the important property that for fixed $(u, v)$

$$\lim_{\sqrt{h^2+k^2}\to 0} \frac{\boldsymbol{\varepsilon}(u, v; h, k)}{\sqrt{h^2 + k^2}} = 0\,.$$

That is, the approximation is not only good, it is good even in comparison with the approximation of $\sqrt{h^2 + k^2}$ to 0. In the language of matrices, we can write this relation as

$$\begin{pmatrix} x(u + h, v + k) \\ y(u + h, v + k) \end{pmatrix} - \begin{pmatrix} x(u, v) \\ y(u, v) \end{pmatrix} \approx \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix}\,.$$

Here the partial derivatives in the $2 \times 2$ matrix are to be evaluated at $(u, v)$. This matrix is called the *Jacobian* matrix of the transformation $\boldsymbol{f}(u, v)$, and in a very suggestive notation, is denoted

$$\frac{\partial(x, y)}{\partial(u, v)}\,.$$

The determinant of this matrix, also called the Jacobian, is denoted as follows:

$$\left|\frac{\partial(x, y)}{\partial(u, v)}\right| = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\,.$$

The chain rule asserts that the Jacobian matrix of a composition of two mappings $(x, y) = \boldsymbol{f}(u, v)$ and $(z, w) = \boldsymbol{g}x, y$ is the product of the two Jacobians, that is,

$$\begin{pmatrix} \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \\ \frac{\partial w}{\partial u} & \frac{\partial w}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}\,.$$

It then follows from linear algebra that

$$\left|\frac{\partial(z,w)}{\partial(u,v)}\right| = \left|\frac{\partial(z,w)}{\partial(x,y)}\right| \left|\frac{\partial(x,y)}{\partial(u,v)}\right| .$$

We assume the reader knows from advanced calculus/elementary real analysis that when the variables are changed in a double integral, the integral transforms using the Jacobian determinant:[3]

$$\iint \Phi(x,y)\,dx\,dy = \iint \Phi\big(x(u,v),y(u,v)\big)\left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv\,.$$

**2.1. The tangent plane.** Having now established our notation, we imagine two observers $A$ and $B$ studying the same surface $\Sigma$ in $\mathbb{R}^3$. For $A$, this surface is given by a parametrization $(u,v) \mapsto \boldsymbol{r}(u,v)$ defined on an open set $U$ in $\mathbb{R}^2$, while for $B$ the parametrization is $(x,y) \mapsto \boldsymbol{s}(x,y)$, defined on an open set $V$ in $\mathbb{R}^2$. The ambient space $\mathbb{R}^3$ is the same for both (the absolute Euclidean space $\mathbb{R}^3$), so that they know when they are talking about the same point $P$ on $\Sigma$, and can communicate the parameter values they give it, thereby establishing a correspondence $(u,v) \leftrightarrow (x,y)$. This correspondence is assumed to be continuously differentiable in both directions, so that the two Jacobian matrices, which are inverses of each other, are defined. Thus, we can regard each set of parameters as functions of the others, and we have the fundamental relation

$$\begin{aligned} x\big(u(s,t),v(s,t)\big) &= s\,, \\ y\big(u(s,t),v(s,t)\big) &= t\,, \\ u\big(x(z,w),y(z,w)\big) &= z\,, \\ v\big(x(z,w),y(z,w)\big) &= w\,. \end{aligned}$$

for all $(s,t) \in V$ and all $(z,w) \in U$.

These equations can be abbreviated as

$$\begin{aligned} \boldsymbol{s}\big(x(u,v),y(u,v)\big) &= \boldsymbol{r}(u,v)\,, \\ \boldsymbol{r}\big(u(x,y),v(x,y)\big) &= \boldsymbol{s}(x,y)\,. \end{aligned}$$

The two observers define the tangent plane to $\Sigma$ at point $P$, to which they assign the respective parameter values $(u_0,v_0)$ and $(x_0,y_0)$, by taking as a basis the partial derivatives

$$\frac{\partial \boldsymbol{r}}{\partial u}\,, \quad \frac{\partial \boldsymbol{r}}{\partial v} \quad \text{and} \quad \frac{\partial \boldsymbol{s}}{\partial x}\,, \frac{\partial \boldsymbol{s}}{\partial y}$$

respectively, evaluated at $(u_0,v_0)$ and $(x_0,y_0)$ and attached to the surface at $P$.

Since these vectors are defined in terms of parameters that $A$ and $B$ do not share, we first ask ourselves if they have defined the same tangent plane. The chain rule provides the answer: The basis vectors being used by each are linear combinations of the basis vectors being used by the other, and hence the tangent space each has defined is contained in the tangent space the other has defined.

[3] We admit only invertible transformations that are defined on connected sets and whose inverses are differentiable. For these, the Jacobian determinant does not vanish and hence is of constant sign. If we want the integral of a positive function to be positive, we need to use the absolute value of the Jacobian instead of the Jacobian itself. This small complication arises when we consider non-oriented integrals.

That is, the tangent space has a meaning independent of the parametrization used to define it:

$$\begin{aligned}
\frac{\partial \boldsymbol{s}(x,y)}{\partial x} &= \frac{\partial u}{\partial x}\frac{\partial \boldsymbol{r}\big(u(x,y),v(x,y)\big)}{\partial u} + \frac{\partial v}{\partial x}\frac{\partial \boldsymbol{r}\big(u(x,y),v(x,y)\big)}{\partial v}\,,\\
\frac{\partial \boldsymbol{s}(x,y)}{\partial y} &= \frac{\partial u}{\partial y}\frac{\partial \boldsymbol{r}\big(u(x,y),v(x,y)\big)}{\partial u} + \frac{\partial v}{\partial y}\frac{\partial \boldsymbol{r}\big(u(x,y),v(x,y)\big)}{\partial v}\,,\\
\frac{\partial \boldsymbol{r}(u,v)}{\partial u} &= \frac{\partial x}{\partial u}\frac{\partial \boldsymbol{s}\big(x(u,v),y(u,v)\big)}{\partial x} + \frac{\partial y}{\partial u}\frac{\partial \boldsymbol{s}\big(x(u,v),y(u,v)\big)}{\partial y}\,,\\
\frac{\partial \boldsymbol{r}(u,v)}{\partial v} &= \frac{\partial x}{\partial v}\frac{\partial \boldsymbol{s}\big(x(u,v),y(u,v)\big)}{\partial x} + \frac{\partial v}{\partial y}\frac{\partial \boldsymbol{s}\big(x(u,v),y(u,v)\big)}{\partial y}\,.
\end{aligned}$$

**2.2. The first fundamental form: Arc length and area.** From their basis tangent vectors, the two observers will each compute the metric coefficients, say $E$, $F$, and $G$ for $A$ and $P$, $Q$, $R$ for $B$. How are these related? We give the coefficients computed by $A$ in terms of those computed by $B$, leaving the easy verification to the reader (Problem 6.2 below):

$$\begin{aligned}
E &= \frac{\partial \boldsymbol{r}}{\partial u}\cdot\frac{\partial \boldsymbol{r}}{\partial u} = \Big(\frac{\partial \boldsymbol{s}}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial \boldsymbol{s}}{\partial y}\frac{\partial y}{\partial u}\Big)\cdot\Big(\frac{\partial \boldsymbol{s}}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial \boldsymbol{s}}{\partial y}\frac{\partial y}{\partial u}\Big)\\
&= P\Big(\frac{\partial x}{\partial u}\Big)^2 + 2Q\,\frac{\partial x}{\partial u}\frac{\partial y}{\partial u} + R\Big(\frac{\partial y}{\partial u}\Big)^2\,,\\
F &= \frac{\partial \boldsymbol{r}}{\partial u}\cdot\frac{\partial \boldsymbol{r}}{\partial v} = \Big(\frac{\partial \boldsymbol{s}}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial \boldsymbol{s}}{\partial y}\frac{\partial y}{\partial u}\Big)\cdot\Big(\frac{\partial \boldsymbol{s}}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial \boldsymbol{s}}{\partial y}\frac{\partial y}{\partial v}\Big)\\
&= P\,\frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + Q\Big(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\Big) + R\,\frac{\partial y}{\partial u}\frac{\partial y}{\partial v}\,,\\
G &= \frac{\partial \boldsymbol{r}}{\partial v}\cdot\frac{\partial \boldsymbol{r}}{\partial v} = \Big(\frac{\partial \boldsymbol{s}}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial \boldsymbol{s}}{\partial y}\frac{\partial y}{\partial v}\Big)\cdot\Big(\frac{\partial \boldsymbol{s}}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial \boldsymbol{s}}{\partial y}\frac{\partial y}{\partial v}\Big)\\
&= P\Big(\frac{\partial x}{\partial v}\Big)^2 + 2Q\,\frac{\partial x}{\partial v}\frac{\partial y}{\partial v} + R\Big(\frac{\partial y}{\partial v}\Big)^2\,.
\end{aligned}$$

The chain rule provides the differential relations

$$\begin{aligned}
du &= \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy\,,\\
dv &= \frac{\partial v}{\partial x}\,dx + \frac{\partial v}{\partial y}\,dy\,.
\end{aligned}$$

When these are combined with the relations for the three metric coefficients, the result, whose rather tedious verification is left to the reader (Problem 6.3) is

$$E\,du^2 + 2F\,du\,dv + G\,dv^2 = P\,dx^2 + 2Q\,dx\,dy + R\,dy^2\,.$$

The interpretation of this differential relation is that if $u$ and $v$ are functions of any variable $t$, then for all values of $t$,

$$\begin{aligned} E\big(u(t),v(t)\big)\big(u'(t)\big)^2 + 2F\big(u(t),v(t)\big)\,u'(t)\,v'(t) + G\big(u(t),v(t)\big)\big(v'(t)\big)^2 = \\ = P\big(x\big(u(t),v(t)\big),y\big(u(t),v(t)\big)\big)\Big(\frac{d}{dt}x\big(u(t),v(t)\big)\Big)^2 + \\ + 2Q\big(x\big(u(t),v(t)\big),y\big(u(t),v(t)\big)\big)\,\frac{d}{dt}x\big(u(t),v(t)\big)\,\frac{d}{dt}y\big(u(t),v(t)\big) + \\ + \big(x\big(u(t),v(t)\big),y\big(u(t),v(t)\big)\big)\Big(\frac{d}{dt}y\big(x\big(u(t),v(t)\big)\big)\Big)^2 . \end{aligned}$$

Thus, as differential forms, the two are identical. Our observers will use these forms to compute arc length and surface area on $\Sigma$. Will they agree on the length of a path and the area of a patch of the surface? Let us see. Suppose $\gamma : [a,b] \to \Sigma$ is a smooth curve on $\Sigma$. The observers $A$ and $B$ will interpret the point $\gamma(t)$ as points $\big(u(t),v(t)\big)$ and $\big(x(t),y(t)\big)$ in $U$ and $V$ respectively, and they will assign to this path the respective lengths

$$\int_a^b \sqrt{E\big(u(t),v(t)\big)\big(u'(t)\big)^2 + 2F\big(u(t),v(t)\big)\,u'(t)v'(t) + G\big(u(t),v(t)\big)}\,dt$$

and

$$\int_a^b \sqrt{P\big(x(t),y(t)\big)\big(x'(t)\big)^2 + 2Q\big(x(t),y(t)\big)\,x'(t)y'(t) + R\big(x(t),y(t)\big)}\,dt\,.$$

Now, precisely because $\boldsymbol{s}\big(x(t),y(t)\big) = \boldsymbol{r}\big(u(t),v(t)\big)$, it follows from the formal identity of the first fundamental form for the two observers (as just written out in detail above), that the integrands in these two expressions are equal at every value of $t$. Hence, they yield the same length for the path $\gamma$.

Similarly, one can show that the two fundamental forms yield the same numerical area for any measurable portion of the surface $\Sigma$. It is again slightly tedious to do so, but one can verify (Problem 6.5) the relation

$$EG - F^2 = \Big|\frac{\partial(x,y)}{\partial(u,v)}\Big|^2 (PR - Q^2)\,.$$

It follows that if $\Sigma_0$ is a portion of $\Sigma$ corresponding to the parameter domains $U_0$ and $V_0$ for the two observers, the area of $\Sigma_0$ is computed as

$$\begin{aligned} \iint\limits_{U_0} \sqrt{EG-F^2}\,du\,dv &= \iint\limits_{U_0} \sqrt{PR-Q^2}\Big|\frac{\partial(x,y)}{\partial(u,v)}\Big|\,du\,dv \\ &= \iint\limits_{V_0} \sqrt{PR-Q^2}\Big|\frac{\partial(x,y)}{\partial(u,v)}\Big|\,\Big|\frac{\partial(u,v)}{\partial(x,y)}\Big|\,dx\,dy \\ &= \iint\limits_{V_0} \sqrt{PR-Q^2}\,dx\,dy\,. \end{aligned}$$

These last two results are the basis for the free use of parametrizations to algebraize the study of surfaces in three-dimensional Euclidean space. We are assured that when we compute length and area, the result of the computation will be the same in any admissible system of parameters, and so we are free to choose the parameters that make the work easiest.

**2.3. Curvature.** We come now to the last of the quantities that we have defined in terms of parameters—the curvature—and we wish to show that its value is also independent of the parametrization. Following the method of Gauss, we allow our two observers to compute the normal vector to the surface as the cross product of the two basis vectors of the tangent space. Once again, the chain rule comes to our aid and shows us that

$$\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v} = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| \left(\frac{\partial \boldsymbol{s}}{\partial x} \times \frac{\partial \boldsymbol{s}}{\partial y}\right).$$

Consequently, $A$ and $B$ compute respectively the unit normal vectors

$$\boldsymbol{m} = \frac{\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}}{\left|\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}\right|} = \frac{\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}}{\sqrt{EG - F^2}}$$

and

$$\boldsymbol{n} = \frac{\frac{\partial \boldsymbol{s}}{\partial x} \times \frac{\partial \boldsymbol{s}}{\partial y}}{\left|\frac{\partial \boldsymbol{s}}{\partial x} \times \frac{\partial \boldsymbol{s}}{\partial y}\right|} = \frac{\frac{\partial \boldsymbol{s}}{\partial x} \times \frac{\partial \boldsymbol{s}}{\partial y}}{\sqrt{PQ - R^2}}.$$

Considering the relation just written for the normal vector and the relation $PQ - R^2 = \left|\frac{\partial(u,v)}{\partial(x,y)}\right|^2 (EG - F^2)$, we see that

$$\boldsymbol{n} = \operatorname{sgn}\left|\frac{\partial(u,v)}{\partial(x,y)}\right| \boldsymbol{m}.$$

But $\operatorname{sgn}\left|\frac{\partial(u,v)}{\partial(x,y)}\right|$ is constant on any connected set, since the Jacobian determinant is non-zero and continuous. Hence, we see that if $A$ computes the expression $DD'' - (D')^2$ that occurs in the numerator of the formula for the curvature, $B$ will compute that this quantity is, with obvious notation,

$$SS'' - (S')^2 = \left|\frac{\partial(u,v)}{\partial(x,y)}\right|^2 \left(DD'' - (D')^2\right).$$

It follows that $A$ and $B$ will compute the curvature to be respectively

$$\pm\sqrt{\frac{DD'' - (D')^2}{EG - F^2}} \text{ and } \pm\sqrt{\frac{SS'' - (S')^2}{PR - Q^2}}.$$

Since $PR - Q^2 = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| (EG - F^2)$, the quantities inside the square roots are equal. The only remaining question is whether the two observers will both choose the ambiguous sign in front of the square root in the same way, that is, whether they will agree as to the curvature being positive or negative.

They will agree. For, as we saw, the ambiguous sign will be chosen by $A$ as the sign of

$$\boldsymbol{m} \cdot \frac{\partial \boldsymbol{m}}{\partial u} \times \frac{\partial \boldsymbol{m}}{\partial v} = \operatorname{sgn}\left|\frac{\partial(x,y)}{\partial(u,v}\right| \boldsymbol{n} \cdot \left|\frac{\partial(x,y)}{\partial(u,v)}\right| \left(\frac{\partial \boldsymbol{n}}{\partial x} \times \frac{\partial \boldsymbol{n}}{\partial y}\right) = \left|\left|\frac{\partial(x,y)}{\partial(u,v)}\right|\right| \left(\boldsymbol{n} \cdot \left(\frac{\partial \boldsymbol{n}}{\partial x} \times \frac{\partial \boldsymbol{n}}{\partial y}\right)\right).$$

Since obviously $\left|\left|\frac{\partial(x,y)}{\partial(u,v)}\right|\right| > 0$, it follows that the two vector triple products

$$\boldsymbol{m} \cdot \frac{\partial \boldsymbol{m}}{\partial u} \times \frac{\partial \boldsymbol{m}}{\partial v} \text{ and } \boldsymbol{n} \cdot \frac{\partial \boldsymbol{n}}{\partial x} \times \frac{\partial \boldsymbol{n}}{\partial y}$$

have the same sign.

That is, $A$ and $B$ will choose the same sign for the curvature. Thus, we have now shown that path length, surface area, and curvature of a surface have a meaning

independent of any particular parametrization in which they are computed. They are intrinsic to the surface.

In order to get to these results, we had to make heavy use of the ambient space $\mathbb{R}^3$ which is an absolute "frame of reference" for both observers. Our quest is to find definitions of length, area, and curvature that are intrinsic to the surface and can be computed without invoking anything "outside" the surface. And, of course, we want to be able to generalize what we do to any manifold, of any dimension, independently of any embedding in a Euclidean space. That is the task we set ourselves in the next section. To tip our hand at the outset, we note that the basic requirement is a set of "metric coefficients" that can be specified as an arbitrary starting point. These are the coefficients $E$, $F$, and $G$ in the discussion above. In practice, they will often come from an absolute manifold embedded in an absolute Euclidean space. But—and this is the important point—it is not necessary to say where they came from. We can use them as the *starting point* for differential geometry. And indeed, in Chapter 4, we did begin by assuming only a general form for these metric coefficients, which we then computed as explicit functions based on an assumed law of gravity (the vanishing of all components of the Ricci tensor).

## 3. Tensors

We now move from the concrete example of surfaces in $\mathbb{R}^3$ to the abstract case of an $n$-dimensional manifold $\mathfrak{M}$. We imagine our two observers $A$ and $B$ "looking" at this manifold by covering it with open sets $U$, on which there are continuous one-to-one mappings $\boldsymbol{\xi} : U \to V_x \subseteq \mathbb{R}^n$ used by $A$ and $\boldsymbol{\eta} : U \to V_y \subseteq \mathbb{R}^n$ used by $B$. To make $\mathfrak{M}$ a $C^\infty$ manifold, we assume that the compositions $\boldsymbol{\xi} \circ \boldsymbol{\eta}^{-1} : V_y \to V_x$ are all $C^\infty$ mappings at each point where they are defined. We shall generally denote these mappings by the coordinates of points in their ranges. Thus we write

$$(y^1, \ldots, y^n) = \boldsymbol{\eta} \circ \boldsymbol{\xi}^{-1}(x^1, \ldots, x^n), \quad (x^1, \ldots, x^n) = \boldsymbol{\xi} \circ \boldsymbol{\eta}^{-1}(y^1, \ldots, y^n),$$

and we note that the Jacobian matrices of these two mappings are inverses of each other:

$$\left|\frac{\partial(y^1, \ldots, y^n)}{\partial(x^1, \ldots, x^n)}\right| \left|\frac{\partial(x^1, \ldots, x^n)}{\partial(y^1, \ldots, y^n)}\right| =$$

$$\begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2} & \cdots & \frac{\partial y^1}{\partial x^{n-1}} & \frac{\partial y^1}{\partial x^n} \\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2} & \cdots & \frac{\partial y^2}{\partial x^{n-1}} & \frac{\partial y^2}{\partial x^n} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \frac{\partial y^{n-1}}{\partial x^1} & \frac{\partial y^{n-1}}{\partial x^2} & \cdots & \frac{\partial y^{n-1}}{\partial x^{n-1}} & \frac{\partial y^{n-1}}{\partial x^n} \\ \frac{\partial y^n}{\partial x^1} & \frac{\partial y^n}{\partial x^2} & \cdots & \frac{\partial y^n}{\partial x^{n-1}} & \frac{\partial y^n}{\partial x^n} \end{pmatrix} \begin{pmatrix} \frac{\partial x^1}{\partial y^1} & \frac{\partial x^1}{\partial y^2} & \cdots & \frac{\partial x^1}{\partial y^{n-1}} & \frac{\partial x^1}{\partial y^n} \\ \frac{\partial x^2}{\partial y^1} & \frac{\partial x^2}{\partial y^2} & \cdots & \frac{\partial x^2}{\partial y^{n-1}} & \frac{\partial x^2}{\partial y^n} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \frac{\partial x^{n-1}}{\partial y^1} & \frac{\partial x^{n-1}}{\partial y^2} & \cdots & \frac{\partial x^{n-1}}{\partial y^{n-1}} & \frac{\partial x^{n-1}}{\partial y^n} \\ \frac{\partial x^n}{\partial y^1} & \frac{\partial x^n}{\partial y^2} & \cdots & \frac{\partial x^n}{\partial y^{n-1}} & \frac{\partial x^n}{\partial y^n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

The information in this matrix equation can be succinctly stated using the Einstein summation convention:

$$\frac{\partial y^i}{\partial x^j}\frac{\partial x^j}{\partial y^k} = \frac{\partial y^i}{\partial y^k} = \delta^i_k = \begin{cases} 1 & \text{if } i = k\,, \\ 0 & \text{if } i \neq k\,. \end{cases} \tag{12}$$

A $C^\infty$ function $f : U \to \mathbb{R}$, defined on an open set $U \subseteq \mathfrak{M}$, will be interpreted by $A$ as $f \circ \boldsymbol{\xi}^{-1} = f_x(x^1, \ldots, x^n)$ and by $B$ as $f \circ \boldsymbol{\eta}^{-1} = f_y(y^1, \ldots, y^n)$. To save ourselves some writing, we shall simply use the letter $f$ to denote both $f_x$ and $f_y$, relying on context to determine which of these functions is meant. Actually, the function $f_y$ is related to the function $f_x$ by the equation

$$f_y(y^1, \ldots, y^n) = f_x\big(\boldsymbol{\xi}\big(\boldsymbol{\eta}^{-1}(y^1, \ldots, y^n)\big)\big)\,.$$

But since this conversion is easily made, we are going to write $\tilde{f}(x^1, \ldots, x^n)$ or $\tilde{f}(y^1, \ldots, y^n)$, both of these expressions standing in for $f(P)$ when $(x^1, \ldots, x^n) = \boldsymbol{\xi}(P)$ and—see the "Feller principle" below—$(y^1, \ldots, y^n) = \boldsymbol{\eta}(P)$. When we discuss vector fields and covector fields, as we are about to do, we shall find that this correspondence allows the chain rule to determine the correct conversions when parameters are changed.

**3.1. Vectors and covectors.** In this abstract setting, a (tangent) vector field $\boldsymbol{u}$ defined on an open set $U \subseteq \mathfrak{M}$ is an operator acting on the set of $C^\infty$-functions on $U$. (And, just as a reminder, a function $f : U \to \mathbb{R}$ is a $C^\infty$-function if $f_x(x^1, \ldots, x^n) = f \circ \boldsymbol{\xi}^{-1}(x^1, \ldots, x^n)$ is a $C^\infty$ function of the parameters $(x^1, \ldots, x^n)$ for some—which implies every—coordinate mapping $\boldsymbol{\xi} : U \to \mathbb{R}^n$.) Like these functions, a vector field is an absolute object, but one kept in the background in most discussions, since we really don't "have" the object, only its representations in terms of different sets of parameters. What makes it a tangent vector field is the structural property $\boldsymbol{u}(fg) = f\boldsymbol{u}(g) + g\boldsymbol{u}(f)$. It is shown in Appendix 4 that $\boldsymbol{u}$ has a representation in terms of the parameters $(x^1, \ldots, x^n)$ in the form

$$u^i \frac{\partial}{\partial x^i} : C^\infty(U) \to C^\infty(U)$$

(here, $u^i$ are $C^\infty$ functions of the parameters), whose effect on a $C^\infty$ function $f : U \to \mathbb{R}$ is given by

$$\boldsymbol{u}(f) = u^i \frac{\partial f \circ \boldsymbol{\xi}^{-1}}{\partial x^i} = u^i \frac{\partial \tilde{f}}{\partial x^i}\,.$$

In writing this equation we are invoking what we referred to above as the *Feller principle*.[4] According to the Feller principle, even though, strictly speaking, the left-hand side of this equation denotes an absolute object and the right-hand side merely its representation in terms of a particular set of parameters, since we don't actually "have" the object, we allow the particular representation to stand in its

[4] As explained by the late William Feller (1906–1970) in a course in probability theory taken by the author in the autumn of 1963, "Ambiguity is what makes mathematics possible." Feller made this remark in the context of explaining that two functions equal to each other almost everywhere are to be regarded as the same function, and more generally, any two objects that satisfy an equivalence relation with each other should be thought of as identical. Mathematicians have been using this principle for centuries, ever since the first mathematician wrote the chain rule as $\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$. The function $z$ is not the same on the two sides of this expression, but we pretend that it is. Without the Feller principle, mathematical papers would be even longer and more tedious than many of them already are. Mathematicians generally invoke this principle by referring to an "abuse of language."

place. Notice that the value of $\boldsymbol{u}(f)$ at a point $Q$ depends on the values of the coefficients $u^i$ *at* the point $\boldsymbol{\xi}(Q)$ and the values of the function $f$ *in a neighborhood of* $\boldsymbol{\xi}(Q)$. It is thus not exactly a "pointwise" object, but rather a "local" object.

In order to discuss the correct change-of-parameter formulas for vector fields, we shall introduce a temporary notation, taking $(u^1, \dots, u^n)$ as the components of a vector field $\boldsymbol{u}$ in $x$-coordinates, we shall denote the components same vector field in $y$ coordinates by $(\tilde{u}^1, \dots, \tilde{u}^n)$. What this means is that for any $C^\infty$ function $f$ on the manifold written as $f_x(x^1, \dots, x^n)$ in $x$-coordinates and $f_y(y^1, \dots, y^n)$ in $y$-coordinates,

$$u^j \frac{\partial f_x}{\partial x^j} = \boldsymbol{u}(f) = \tilde{u}^j \frac{\partial f_y}{\partial y^j}\,.$$

Here again, $\tilde{f}_y(y^1, \dots, y^n) = \tilde{f}_x\big(x^1(y^1, \dots, y^n), \dots, x^n(y^1, \dots, y^n)\big)$, and conversely, $\tilde{f}_x(x^1, \dots, x^n) = \tilde{f}_y\big(y^1(x^1, \dots, x^n), \dots, y^n(x^1, \dots, x^n)\big)$.

By the chain rule, we see that in $y$-parameters

$$\boldsymbol{u}(f) = u^j \frac{\partial y^i}{\partial x^j} \frac{\partial \tilde{f}}{\partial y^i} = \tilde{u}^i \frac{\partial \tilde{f}}{\partial y^i}\,,$$

where (observing the summation convention)

$$\tilde{u}^i = u^j \frac{\partial y^i}{\partial x^j}\,.$$

Again using the Feller principle,[5] we say that, *as tensors*,

$$\boldsymbol{u} = u^i \frac{\partial}{\partial x^i} = \tilde{u}^i \frac{\partial}{\partial y^i}\,, \tag{13}$$

where

$$\tilde{u}^i = u^j \frac{\partial y^i}{\partial x^j}\,.$$

This last equality holds when the corresponding parameters are used on both sides, and only on parameter domains that are the images of the same open subset of $\mathfrak{M}$.

As remarked in Appendix 4, this transformation law is characteristic of what are called *contravariant* tensors. The "contrariness" of contravariance manifests itself in the relation

$$\frac{\partial}{\partial y^i} = \frac{\partial x^j}{\partial y^i} \frac{\partial}{\partial x^j}\,.$$

Although this equation is thought of as an equation that defines the left-hand side (an expression in $y$-variables) in terms of the right-hand side (an expression in $x$-variables), the Jacobian on the right actually treats the $y^i$ as the independent variables, and thus would normally be written in terms of the $y^i$. It is the Jacobian of the transformation $(y^1, \dots, y^n) \mapsto (x^1, \dots, x^n)$, although the equation itself is used to convert from $x^j$ to $y^i$. The two conversions go in opposite directions; hence the term *contravariant*.

[5] Note in particular that we are using the symbol $\partial/\partial x^i$ to denote a vector, that is, an operator on functions on the manifold, whereas from its original definition, unexpanded by the Feller principle, it operates only on functions on the parameter space. We are setting the absolute object equal to two algebraic expressions representing operators, neither of which can possibly be equal to it as a matter of pure logic, and which also are generally not equal to each other, since they operate on different spaces of functions! "A foolish consistency is the hobgoblin of little minds" (Ralph Waldo Emerson).

In contrast, if we look at differentials, we find the relation

$$dy^i = \frac{\partial y^i}{\partial x^j}\, dx^j \,,$$

and here the conversions both go in the same direction. Hence, a differential is a *covariant* object.

As shown in Appendix 4, at each point $P \in \mathfrak{M}$, the mappings $\boldsymbol{u}_P f = \boldsymbol{u} f\big(\boldsymbol{\xi}(P)\big)$ form a vector space of dimension $n$, the tangent space at $P$. Like all vector spaces, this one has a dual space of the same dimension, called the cotangent space, and for each basis $\{\boldsymbol{u}_i\}$ of the tangent space, there is a unique dual basis $\{\boldsymbol{v}_j^*\}$ of the cotangent space[6] having the property that

$$\boldsymbol{v}_j^*(\boldsymbol{u}_i) = \delta_j^i \,.$$

As shown in courses of linear algebra, when addition and scalar multiplication are defined in the obvious way on the cotangent space, its dual space is naturally identified with the tangent space on which its elements (covectors) act. Because of that symmetry, we usually write[7]

$$\langle \boldsymbol{u}_i, \boldsymbol{v}_j^* \rangle = \delta_i^j$$

when the two sets of vectors and covectors are dual bases, thus allowing vector fields and covector fields to act on each other.

Since obviously covectors transform covariantly under a change of parametrization, we find it convenient to identify the dual basis of the standard basis of the tangent space in a given parametrization, which is $\{\partial/\partial x^i\}$, with the space of differentials $\{dx^j\}$.

The change-of-parameter formula for covector fields can thus be read as the equation (see the Feller principle)

$$b_i\, dx^i = b_i \frac{\partial x^i}{\partial y^j}\, dy^j \,, \tag{14}$$

Only one habit formed in calculus needs to be broken here, and that is the habit of thinking of $dx^i$ as an infinitesimal increment in $x^i$. It is really the partial derivative operator $\partial/\partial x^i$ that stands in for an infinitesimal increment in $x^i$ in this context. The covector $dx^i$ is defined in purely finite, algebraic terms, by the equation

$$\left\langle b_i dx^i, a^j \frac{\partial}{\partial x^j} \right\rangle = b_i a^i \,.$$

---

[6] Here again a certain notational awkwardness arises. We use subscripts to denote the coordinates of covectors and superscripts to denote the coordinates of contravariant vectors. That leaves us "nowhere to go" when we need to talk about an indexed set of vectors and covectors. Our convention in this appendix is to use subscripts to index sets of both types, appending a superscripted star when the objects are covectors.

[7] The angle-bracket notation here requires both a vector and a covector. In Chapter 6, we defined a natural *inner product* on the tangent space alone, and we also used the angle-bracket notation there. In that situation, both arguments were (contravariant) vectors. Recall, though, that we defined the inner product $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ as $g_{ij} u^i v^j$. This can be thought of, in our present notation, as $\langle \boldsymbol{u}^*, \boldsymbol{v} \rangle$, where $\boldsymbol{u}^*$ is the covector obtained by lowering the superscripts that denote the components of the vector $\boldsymbol{u}$, that is, $u_j^* = g_{ij} u^i$. There is thus no ambiguity in our use of the same notation in both situations.

**Remark 6.1.** The interaction of a vector field and a covector field is a *pointwise* function, depending only on the values the cofficients $a^i$ and $b_j$ have *at* a particular point. It does not require any nearby values. (On the other hand, the coefficients of the vector and covector fields we deal with will very often themselves be derivatives and hence *will* depend on nearby values of the functions they are derivatives of.)

Notice also that, while vector fields and covector fields are dual to each other, the symmetry between them is not quite complete, since vector fields were originally defined as local directional derivative operators acting on $C^\infty(U)$. Covector fields were defined so as to be auxiliary to vector fields.

We have now established principles for deciding when two observers are talking about the same point on a manifold, when they are talking about the same function on the manifold, and when they are talking about the same tangent vectors and covectors at a given point. These are the basic objects from which differential geometry will build such concepts as curvature, geodesics, affine connections, and the like. In every case, the issue that needs to be settled before we can have any confidence in the usefulness of our definitions is the following: If Observer $A$ combines geometric objects $a$, $b$, and $c$ to produce an object $d$, and Observer B, to whom these objects are $\tilde{a}$, $\tilde{b}$, and $\tilde{c}$ according to the correspondence rules we have established, combines them in exactly the same way to produce an object $\tilde{d}$, *will* $\tilde{d}$ *correspond to* $d$ *under the rules we have given?* Since differential geometry is based on the effective application of differentiation, all our correspondence rules involve Jacobian matrices. Thus, we are asking whether, when the "ingredients" are translated from $x$-coordinates to $y$-coordinates using rows or columns of the Jacobian matrix, will the final product also be translated correctly using the Jacobian matrix? If the answer is positive, the object in question is called a *tensor*. A more precise definition now follows.

**3.2. Multilinear functions. The metric coefficients.** A scalar-valued function defined on an open set $U \subseteq \mathfrak{M}$ will sometimes be referred to as a *tensor field of type* $(0,0)$. The formula for changing parameters in such a function is obvious and trivial: The number returned as the value of the function at each point of $U$ must be *the same real number*, that is, if $(x^1, \dots x^n)$ and $(y^1, \dots, y^n)$ are the coordinates of the same point $Q$ in two different systems, a function $f(Q)$ interpreted as $f_x(x^1, \dots, x^n)$ and $f_y(y^1, \dots, y^n)$ in the two systems, must be such that $f_x(x^1, \dots, x^n) = f_y(y^1, \dots, y^n)$ as real numbers. Only then are we justified in using the notation $f(Q)$ and saying that $f$ is a tensor field of type $(0,0)$. We have seen two examples of such tensors above, in the length of a fixed path through a point $P$ from $P$ to a variable point $Q$ on the path, and in the curvature of a surface at the point $Q$.

A vector field will be referred to as a tensor field of type $(1,0)$ and a covector field as a tensor field of type $(0,1)$. Since we can regard vector fields as acting on covector fields, the way is now clear to discuss multilinear functions of all types. Consider, for example, a surface parameterized by $(u,v) \mapsto \boldsymbol{\rho}(u,v)$ and $(s,t) \mapsto \boldsymbol{\sigma}(s,t)$. The first fundamental form on this surface is expressed in the usual way, using dot products of the partial derivatives of $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$, and we can switch

parameters via the formulas

$$\begin{aligned}
\widetilde{E}(s,t) &= \frac{\partial \boldsymbol{\sigma}}{\partial s}\cdot\frac{\partial \boldsymbol{\sigma}}{\partial s} = \Big(\frac{\partial u}{\partial s}\frac{\partial \boldsymbol{\rho}}{\partial u}+\frac{\partial v}{\partial s}\frac{\partial \boldsymbol{\rho}}{\partial v}\Big)\cdot\Big(\frac{\partial u}{\partial s}\frac{\partial \boldsymbol{\rho}}{\partial u}+\frac{\partial v}{\partial s}\frac{\partial \boldsymbol{\rho}}{\partial v}\Big)\\
&= E(u,v)\Big(\frac{\partial u}{\partial s}\Big)^2+2F(u,v)\,\frac{\partial u}{\partial s}\,\frac{\partial v}{\partial s}+G(u,v)\,\Big(\frac{\partial v}{\partial s}\Big)^2,\\
\widetilde{F}(s,t) &= E(u,v)\,\frac{\partial u}{\partial s}\,\frac{\partial u}{\partial t}+F(u,v)\,\Big(\frac{\partial u}{\partial s}\,\frac{\partial v}{\partial t}+\frac{\partial u}{\partial t}\,\frac{\partial v}{\partial s}\Big)+G(u,v)\,\frac{\partial v}{\partial s}\,\frac{\partial v}{\partial t},\\
\widetilde{G}(s,t) &= E(u,v)\Big(\frac{\partial u}{\partial t}\Big)^2+2F(u,v)\,\frac{\partial u}{\partial t}\,\frac{\partial v}{\partial t}+G(u,v)\,\Big(\frac{\partial v}{\partial t}\Big)^2.
\end{aligned}$$

If we write $g_{11}=E$, $g_{12}=g_{21}=F$, $g_{22}=G$, $\tilde{g}_{11}=\widetilde{E}$, $\tilde{g}_{12}=\tilde{g}_{21}=\widetilde{F}$, and $\tilde{g}_{22}=\widetilde{G}$, with and let $x^1=u$, $x^2=v$, $y^1=s$, $y^2=t$, we can write these formulas using the summation convention as

$$\tilde{g}_{ij}=g_{kl}\,\frac{\partial x^k}{\partial y^i}\,\frac{\partial x^l}{\partial y^j}\,. \tag{15}$$

We need the metric coefficients $g_{ij}$ to define arc length $L$ along a parameterized curve $C$ given in $x$-coordinate system as the image of an interval $a\le t\le b$ under a mapping $t\mapsto \boldsymbol{r}\big(x^1(t),x^2(t)\big)$:

$$L=\int_a^b\sqrt{g_{ij}(x^i)'(t)(x^j)'(t)}\,dt\,.$$

We can regard the metric coefficients $g_{ij}$ as the coefficients of a bilinear mapping $(\boldsymbol{u},\boldsymbol{v})\mapsto\mathbb{R}$. If $\boldsymbol{u}=u^i(\partial)/(\partial x^i)$ and $\boldsymbol{v}=v^j(\partial)/(\partial x^j)$, this mapping is computed in the $x$-coordinate system as the natural inner product of two contravariant vectors.

$$g_{ij}u^iv^j=\langle\boldsymbol{u},\boldsymbol{v}\rangle\,.$$

**3.3. Tensor products.** Yet another way of looking at this last equation involves what is called the *tensor product* of two tensors. The basis of the cotangent space dual to the basis $\{\partial/\partial x^i\}$ is $dx^i$, where by definition $dx^i\big(u^j(\partial/\partial x^j)\big)=u^i$. The covector $dx^i$ is a linear functional that operates on a single contravariant vector $u^j\partial/\partial x^j$ to produce the real number $u^i$. As already mentioned, it is also referred to as a *tensor of type* $(0,1)$. Conversely, the partial derivative $\partial/\partial x^j$ can be regarded as a linear functional operating on a single covector $u_i\,dx^i$ ot produce the real number $u_i$, and so a contravariant vector can be called a *tensor of type* $(1,0)$.

We then define the *tensor product* $dx^i\otimes dx^j$ as the bilinear operator acting on a pair of contravariant vectors $u^k\,\partial/\partial x^k$ and $v^l\,\partial/\partial x^l$ and defined as the product

$$dx^i\otimes dx^j\Big(u^k\frac{\partial}{\partial x^k},v^l\frac{\partial}{\partial x^l}\Big)=dx^i\Big(u^k\frac{\partial}{\partial x^k}\Big)dx^j\Big(v^l\frac{\partial}{\partial x^l}\Big)=u^iv^j\,.$$

In this way, the metric coefficients become the coordinates of a *tensor of type* $(0,2)$, namely the tensor

$$g_{ij}dx^i\otimes dx^j\,.$$

Equation (15) describes the transformation of the first fundamental form on $n$-dimensional manifolds. As long as we are discussing the rules for conversion from one set of parameters to another, it is useful to retain the tilde notation to show how the coefficients must be transformed, and we shall continue to do that until we have discussed all the cases we need to discuss. Once that labor is finished, we

really can omit the tilde, since the differentials $dx^i$ or $dy^i$ or the partial derivatives $\partial/\partial x^i$ or $\partial/\partial y^i$ suffice to indicate which parameters are being used.

To recapitulate: The first fundamental form is a bilinear mapping $\lambda(\boldsymbol{u},\boldsymbol{v})$ that takes a pair of tangent vector fields $\boldsymbol{u} = u^i\,\partial/\partial x^i$ and $\boldsymbol{v} = v^j\,\partial/\partial x^j$ to the scalar function

$$g_{ij}u^iv^j = g_{ij}\,dx^i(\boldsymbol{u})\,dx^j(\boldsymbol{v}) = \lambda(\boldsymbol{u},\boldsymbol{v})\,.$$

We call the first fundamental form a *tensor of type* $(0,2)$. Its action produces the *standard inner product* on the tangent space, written $\langle\boldsymbol{u},\boldsymbol{v}\rangle$. Thus

$$\langle\boldsymbol{u},\boldsymbol{v}\rangle = \lambda(\boldsymbol{u},\boldsymbol{v}) = g_{ij}u^iv^j\,.$$

**Remark 6.2.** If $g^{ij}$ denotes the $(i,j)$th entry of the inverse of the matrix whose $(i,j)$th entry is $g_{ij}$, it is easy to show that the $g^{ij}$ are the coordinates of a tensor of type $(2,0)$. For, consider covector fields $\boldsymbol{u}^* = u_i\,dx^i = u_i(\partial x^i/\partial y^k)\,dy^k = \tilde{u}_k\,dy^k$ and $\boldsymbol{v}^* = v_j\,dx^j = v_j(\partial x^j/\partial y^l)\,dy^l = \tilde{v}_l\,dy^l$. Then

$$\tilde{g}^{kl}\tilde{u}_k\tilde{v}_l = \tilde{g}^{kl}u_iv_j\frac{\partial x^i}{\partial y^k}\,\frac{\partial x^j}{\partial y^l} = \hat{g}^{ij}u_iv_j\,,$$

where

$$\hat{g}^{ij} = \tilde{g}^{kl}\frac{\partial x^i}{\partial y^k}\,\frac{\partial x^j}{\partial y^l}\,.$$

Now the functions $\tilde{g}^{kl}$ are the entries in the inverse of the matrix whose entries are $\tilde{g}_{mn}$, and

$$g_{pq} = \tilde{g}_{mn}\frac{\partial y^m}{\partial x^p}\,\frac{\partial y^n}{\partial x^q}\,.$$

Thus, we have

$$\begin{aligned} g_{pq}\hat{g}^{ij} &= g_{pq}\tilde{g}^{kl}\frac{\partial x^i}{\partial y^k}\,\frac{\partial x^j}{\partial y^l} \\ &= \tilde{g}^{kl}\tilde{g}_{mn}\frac{\partial y^m}{\partial x^p}\,\frac{\partial y^n}{\partial x^q}\frac{\partial x^i}{\partial y^k}\,\frac{\partial x^j}{\partial y^l}\,. \end{aligned}$$

In this last expression, we set $q = i$ and sum over $i$ (using the summation convention). We find that

$$\begin{aligned} g_{pi}\hat{g}^{ij} &= \tilde{g}^{kl}\tilde{g}_{mn}\frac{\partial y^m}{\partial x^p}\,\frac{\partial y^n}{\partial x^i}\frac{\partial x^i}{\partial y^k}\,\frac{\partial x^j}{\partial y^l} \\ &= \delta^n_k\tilde{g}^{kl}\tilde{g}_{mn}\frac{\partial y^m}{\partial x^p}\,\frac{\partial x^j}{\partial y^l} \\ &= \tilde{g}^{kl}\tilde{g}_{mk}\frac{\partial y^m}{\partial x^p}\,\frac{\partial x^j}{\partial y^l} \\ &= \delta^l_m\frac{\partial y^m}{\partial x^p}\,\frac{\partial x^j}{\partial y^l} \\ &= \frac{\partial y^l}{\partial x^p}\,\frac{\partial x^j}{\partial y^l} \\ &= \delta^j_p\,. \end{aligned}$$

This means that the matrix whose entries are $\hat{g}^{ij}$ is the inverse of the matrix whose entries are $g_{ij}$, and thus that $\hat{g}^{ij} = g^{ij}$. In other words, we have proved that

$$\tilde{g}^{ij}\tilde{u}_i\tilde{v}_j = g^{ij}u_iv_j\,,$$

and so we have verified that $g^{kl}$ are the coefficients of a tensor of type $(2,0)$.

**Remark 6.3.** The metric cofficients play a fundamental role in everything we do, as we have seen in Chapters 4–6. They can be used to convert a contravariant vector into a covariant vector:

$$\boldsymbol{u} = u^i \frac{\partial}{\partial x^i} \mapsto \boldsymbol{u}^* = g_{ij} u^i \, dx^j = \boldsymbol{u}^* \, .$$

It is quite trivial to verify that $\boldsymbol{u}^*$ is a covector; that is, that it transforms correctly when we change to $y$-coordinates. This operation, when discussed entirely in terms of coordinates, as the mapping $\{u^i\} \mapsto \{g_{ij}u^i\} = \{v_j\}$, is known as *lowering an index*. The operation can be applied very generally to lower any index of any tensor that has contravariant components. And of course, one can use the inverse matrix to raise an index.

In general, a tensor of type[8] $(k,l)$ (rank $k+l$) is a multilinear mapping $\mu(\boldsymbol{v}_1^*, \ldots, \boldsymbol{v}_k^*; \boldsymbol{u}_1, \ldots, \boldsymbol{u}_l)$ of the Cartesian product of $k$ covector fields $\boldsymbol{v}_1^*, \ldots, \boldsymbol{v}_k^*$ and $l$ vector fields $\boldsymbol{u}_1, \ldots \boldsymbol{u}_l$. It is indexed by a set of scalar-valued functions $T^{j_1,\ldots,j_k}_{i_1,\ldots,i_l}$ of the parameters $x^1, \ldots, x^n$ and acts according to the formula

$$\mu(\boldsymbol{v}_1^*, \ldots, \boldsymbol{v}_k^*; \boldsymbol{u}_1, \ldots, \boldsymbol{u}_l) = T^{j_1,\ldots,j_k}_{i_1,\ldots,i_l} v^*_{1j_1} \cdots v^*_{kj_k} u_1^{i_1} \cdots u_l^{i_l} \, .$$

Of course, given a parametrization, *any* set of $n^{k+l}$ $C^\infty$-functions $T^{j_1,\ldots,j_k}_{i_1,\ldots,i_l}(x^1, \ldots, x^n)$ *will* define such a multilinear operator. But this operator will define a tensor only if it is agreed that when the parameters are changed from $x$ to $y$, the corresponding operator is the one having the coefficients

$$\widetilde{T}^{j_1,\ldots,j_k}_{i_1,\ldots,i_l} = T^{n_1,\ldots,n_k}_{m_1,\ldots,m_l} \frac{\partial x^{m_1}}{\partial y^{i_1}} \cdots \frac{\partial x^{m_l}}{\partial y^{i_l}} \frac{\partial y^{j_1}}{\partial x^{n_1}} \cdots \frac{\partial y^{j_k}}{\partial x^{n_k}} \, . \tag{16}$$

Merely being a multilinear operator in one set of coordinates does not suffice to make a tensor. The crucial thing is the transformation law. In Chapters 5 and 6 we have encountered the metric and Ricci tensors, both of type $(0,2)$, and the Riemann curvature tensor, of type $(1,3)$.

This point is important in geometry, since all the operations involve differentiation, and all conversions must therefore be performed using the entries in the Jacobian matrices of the coordinate change.

The algebra underlying multilinear functions is quite transparent, although the multiple subscripts and superscripts can sometimes be confusing. Many simple facts that need to be proved, facts that would tax even so sophisticated a computer algebra system as *Mathematica*, can be easily derived by using these indices and the summation convention, as we shall now demonstrate.

**3.4. The Lie bracket.** We wish to establish that the Lie bracket of two vector fields $\boldsymbol{u} = u^i(\partial/\partial x^i)$ and $\boldsymbol{v} = v^j(\partial/\partial x^j)$ is a tensor of type $(1,0)$, that is, a contravariant vector. We defined this operation in Chapter 6, and we know already that the mapping $(\boldsymbol{u}, \boldsymbol{v}) \mapsto [\boldsymbol{u}, \boldsymbol{v}] = \boldsymbol{u}\,\boldsymbol{v} - \boldsymbol{v}\,\boldsymbol{u}$ is a bilinear mapping of the Cartesian product of a local "tangent bundle" with itself into itself. As long as we stick with a single parametrization, no problems arise. But we need to see that if $\boldsymbol{u}$ and $\boldsymbol{v}$

[8] Most unfortunately, this notation is not universally used, and what we are calling a tensor of type $(k,l)$ will be called a tensor of type $(l,k)$ in sources that the reader is liable to encounter on the Internet. In particular, some sources refer to the Riemann curvature tensor as a "(3,1) tensor." Try not to let this inconsistency worry you. Most contemporary books adhere to the notation given here, if they bother to distinguish types within the rank $k+l$ at all.

are expressed in different coordinates, the bracket of their new forms is the appropriate transformation of the bracket of the original pair. There is no question that the mapping $(\boldsymbol{u},\boldsymbol{v})\mapsto \boldsymbol{u}\,\boldsymbol{v}-\boldsymbol{v}\,\boldsymbol{u}$ produces a tangent vector (a mapping of $C^\infty(U)$ into itself that has the derivation property $[\boldsymbol{u},\boldsymbol{v}](fg)=f[\boldsymbol{u},\boldsymbol{v}]g+g[\boldsymbol{u},\boldsymbol{v}]f$). What we need to see is that if we change parameters from $x$ to $y$ and recompute the Lie bracket in those new parameters, the result will represent the same absolute tangent vector.

**Theorem 6.1.** *The Lie bracket $[\boldsymbol{u},\boldsymbol{v}]$ of two tangent vector fields is a tangent vector field, that is, a tensor of type $(1,0)$. In terms of coordinates, this means that*

$$\tilde{u}^i\frac{\partial\tilde{v}^j}{\partial y^i}-\tilde{v}^i\frac{\partial\tilde{u}^j}{\partial y^i}=\frac{\partial y^j}{\partial x^l}\Big(u^k\frac{\partial v^l}{\partial x^k}-v^k\frac{\partial u^l}{\partial x^k}\Big)\,.$$

*As vectors, it says that*

$$\Big(\tilde{u}^i\frac{\partial\tilde{v}^j}{\partial y^i}-\tilde{v}^i\frac{\partial\tilde{u}^j}{\partial y^i}\Big)\frac{\partial}{\partial y^j}=\Big(u^i\frac{\partial v^j}{\partial x^i}-v^i\frac{\partial u^j}{\partial x^i}\Big)\frac{\partial}{\partial x^j}\,.$$

Proof. As in all the proofs in this chapter, the procedure is a simple, mindless linking of formulas. The steps are as follows. Assume that $\boldsymbol{u}=u^i\partial/\partial x^i$ and $\boldsymbol{v}=v^j\partial/\partial x^j$ in $x$-coordinates. Then, for any $C^\infty$ function $f$, we have

$$\begin{aligned}\tilde{\boldsymbol{u}}\big(\tilde{\boldsymbol{v}}\tilde{f}\big) &= u^k\frac{\partial y^i}{\partial x^k}\frac{\partial}{\partial y^i}\Big(v^l\frac{\partial y^j}{\partial x^l}\frac{\partial\tilde{f}}{\partial y^j}\Big)\\ &= u^k\frac{\partial y^i}{\partial x^k}\bigg(\frac{\partial v^l}{\partial y^i}\,\frac{\partial y^j}{\partial x^l}\frac{\partial\tilde{f}}{\partial y^j}+v^l\frac{\partial}{\partial y^i}\Big(\frac{\partial y^j}{\partial x^l}\Big)\frac{\partial\tilde{f}}{\partial y^j}+v^l\frac{\partial y^j}{\partial x^l}\,\frac{\partial^2\tilde{f}}{\partial y^i\,\partial y^j}\bigg)\,.\end{aligned}$$

The awkward part here—due to the contravariance of these vectors—is the derivative

$$\frac{\partial}{\partial y^i}\Big(\frac{\partial y^j}{\partial x^l}\Big)=\frac{\partial x^m}{\partial y^i}\,\frac{\partial}{\partial x^m}\Big(\frac{\partial y^j}{\partial x^l}\Big)=\frac{\partial x^m}{\partial y^i}\,\frac{\partial^2 y^j}{\partial x^m\,\partial x^l}\,.$$

The expression $\tilde{\boldsymbol{u}}\big(\tilde{\boldsymbol{v}}\tilde{f}\big)$ therefore contains a term with coefficient

$$\begin{aligned}u^kv^l\frac{\partial y^i}{\partial x^k}\frac{\partial x^m}{\partial y^i}\frac{\partial^2 y^j}{\partial x^m\,\partial x^j} &= u^kv^l\frac{\partial x^m}{\partial x^k}\frac{\partial^2 y^j}{\partial x^m\,\partial x^l}\\ &= u^kv^l\frac{\partial^2 y^j}{\partial x^k\,\partial x^l}\,.\end{aligned}$$

We then have

$$\tilde{\boldsymbol{u}}\big(\tilde{\boldsymbol{v}}\tilde{f}\big)=u^k\frac{\partial y^i}{\partial x^k}\,\frac{\partial v^l}{\partial y^i}\frac{\partial y^j}{\partial x^l}\frac{\partial\tilde{f}}{\partial y^j}+u^kv^l\frac{\partial^2 y^j}{\partial x^k\,\partial x^l}\,\frac{\partial\tilde{f}}{\partial y^j}+u^kv^l\frac{\partial y^i}{\partial x^k}\frac{\partial y^j}{\partial x^l}\frac{\partial^2\tilde{f}}{\partial y^i\,\partial y^j}\,,$$

$$\tilde{\boldsymbol{v}}\big(\tilde{\boldsymbol{u}}\tilde{f}\big)=v^l\frac{\partial y^i}{\partial x^l}\,\frac{\partial u^k}{\partial y^i}\frac{\partial y^j}{\partial x^k}\frac{\partial\tilde{f}}{\partial y^j}+v^lu^k\frac{\partial^2 y^j}{\partial x^l\,\partial x^k}\,\frac{\partial\tilde{f}}{\partial y^j}+v^lu^k\frac{\partial y^i}{\partial x^l}\frac{\partial y^j}{\partial x^k}\frac{\partial^2\tilde{f}}{\partial y^i\,\partial y^j}\,.$$

The second of these relations is obtained from the first by interchanging $k$ with $l$ and $u$ with $v$. The first of these interchanges makes no difference, since we are summing over these indices, which are therefore "dummy variables." The interchange of $u$ with $v$ is necessitated by the interchange of $\boldsymbol{u}$ with $\boldsymbol{v}$.

When we subtract the second equation from the first, we find that all the terms involving second-order partial derivatives cancel out, and, due to the chain

rule applied to the remaining terms,

$$[\tilde{\boldsymbol{u}},\tilde{\boldsymbol{v}}]f = [\boldsymbol{u},\boldsymbol{v}]f = \Big(u^k\frac{\partial v^l}{\partial x^k} - v^k\frac{\partial u^l}{\partial x^k}\Big)\frac{\partial f}{\partial x^l}\,.$$

□

**3.5. What *isn't* a tensor? The Christoffel symbols.** The straightforward, if messy, procedure by which we verified that the first fundamental form, the inverse of its matrix, and the Lie bracket of two contravariant vectors are all tensors can easily lead one to the conclusion that every multilinear expression is a tensor. That such is not the case is best shown by the example of the set of Christoffel symbols $\Gamma^i_{jk}$, which looks suspiciously like a tensor of type $(1,2)$, but isn't. Like so many of the important results in this area, the actual process involves a tedious, but not difficult computation, which we now give, to exhibit the relation between the Christoffel symbol $\widetilde{\Gamma}^i_{jk}$ in $y$-parameters and the Christoffel symbol $\Gamma^p_{qr}$ in $x$-parameters.

In this subsection and in the next, we shall make extensive use of the inverse relation between the Jacobian matrices, that is, Eq. (12).

**Theorem 6.2.** *The Christoffel symbols $\Gamma^i_{jk}$ are* not *a tensor of type $(1,2)$. When coordinates are changed from $x$ to $y$, the relation between the Christoffel symbol $\Gamma^p_{qr}$ in $x$-coordinates and the Christoffel symbol $\widetilde{\Gamma}^i_{jk}$ in $y$-coordinates is*

$$\widetilde{\Gamma}^i_{jk} = \frac{\partial y^i}{\partial x^p}\,\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^r}{\partial y^k}\,\Gamma^p_{qr} + \frac{\partial y^i}{\partial x^p}\,\frac{\partial^2 x^p}{\partial y^j\,\partial y^k}\,.$$

Proof.

$$\begin{aligned}\widetilde{\Gamma}^i_{jk} &= \frac{1}{2}\tilde{g}^{il}\Big(\frac{\partial \tilde{g}_{jl}}{\partial y^k} + \frac{\partial \tilde{g}_{lk}}{\partial y^j} - \frac{\partial \tilde{g}_{jk}}{\partial y^l}\Big)\\ &= \frac{1}{2}g^{ps}\frac{\partial y^i}{\partial x^p}\,\frac{\partial y^l}{\partial x^s}\Big(\frac{\partial}{\partial y^k}\Big(g_{qr}\frac{\partial x^q}{\partial y^j}\frac{\partial x^r}{\partial y^l}\Big) + \frac{\partial}{\partial y^j}\Big(g_{qr}\frac{\partial x^q}{\partial y^k}\frac{\partial x^r}{\partial y^l}\Big)\Big)\\ &\qquad - \frac{1}{2}g^{ps}\frac{\partial y^i}{\partial x^p}\,\frac{\partial y^l}{\partial x^s}\frac{\partial}{\partial y^l}\Big(g_{qr}\frac{\partial x^q}{\partial y^j}\frac{\partial x^r}{\partial y^k}\Big)\,.\end{aligned}$$

Since the first two terms are handled identically, we shall give the details just for the first of them. Considering just the portion of it inside the large parentheses, we make the substitution $\partial/\partial y^k = (\partial x^m/\partial y^k)(\partial/\partial x^m)$, and expand it:

$$\frac{\partial x^m}{\partial y^k}\Big(\frac{\partial g_{qr}}{\partial x^m}\,\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^r}{\partial y^l} + g_{qr}\frac{\partial}{\partial x^m}\Big(\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^r}{\partial y^l}\Big)\Big)\,.$$

We now have two "subterms" to consider. When the first "subterm" is multiplied by the factor $\partial y^l/\partial x^s$, which is outside the large parentheses in the expression for $\widetilde{\Gamma}^i_{jk}$, the result contains a factor

$$\frac{\partial y^l}{\partial x^s}\,\frac{\partial x^r}{\partial y^l} = \delta^r_s\,.$$

We replace $r$ by $s$, throughout, causing $r$ to disappear as an index of summation. Our first "subterm" then becomes, when the outside factors are multiplied through,

$$\frac{1}{2}g^{ps}\frac{\partial y^i}{\partial x^p}\,\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^m}{\partial y^k}\,\frac{\partial g_{qs}}{\partial x^m}\,.$$

Since $r$ has disappeared, we are free to use it in place of $m$ as an index of summation, and we shall find that convenient to do. Thus our first "subterm" is

$$\frac{1}{2}g^{ps}\frac{\partial y^i}{\partial x^p}\,\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^r}{\partial y^k}\,\frac{\partial g_{qs}}{\partial x^r}\,.$$

We shall consider the second "subterm" in the first term after we have finished discussing the first "subterms" in all three of the original terms.

The second term inside the large parentheses expands, when we make the substitution $\partial/\partial y^j = (\partial x^m/\partial y^j)(\partial/\partial x^m)$, to become

$$\frac{\partial x^m}{\partial y^j}\left(\frac{\partial g_{qr}}{\partial x^m}\,\frac{\partial x^q}{\partial y^k}\,\frac{\partial x^r}{\partial y^l} + g_{qr}\frac{\partial}{\partial x^m}\Big(\frac{\partial x^q}{\partial y^k}\,\frac{\partial x^r}{\partial y^l}\Big)\right).$$

Once again, when we multiply through by $\partial y^l/\partial x^s$, we get a factor $\delta^r_s$, which again causes $r$ to disappear as a summation index. If we then replace $m$ with $r$, we get the "subterm"

$$\frac{1}{2}g^{ps}\frac{\partial y^i}{\partial x^p}\,\frac{\partial x^r}{\partial y^j}\,\frac{\partial x^q}{\partial y^k}\frac{\partial g_{qs}}{\partial x^r}\,.$$

Since both $r$ and $q$ are dummy indices of summation, we can interchange them so that this third "subterm" looks more like the first one, namely

$$\frac{1}{2}g^{ps}\frac{\partial y^i}{\partial x^p}\,\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^r}{\partial y^k}\,\frac{\partial g_{rs}}{\partial x^q}\,.$$

Again, leaving the fourth "subterm" (the second subterm of the second term) in abeyance for the moment, let us turn to the fifth "subterm," which is the first "subterm" of the third term. We do not need to replace the partial derivative operator $\partial/\partial y^l$ in this case, since, when we multiply through by $\partial y^l/\partial x^s$, we get immediately the "subterm"

$$-\frac{1}{2}g^{ps}\frac{\partial y^i}{\partial x^p}\,\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^r}{\partial y^k}\frac{\partial g_{rq}}{\partial x^s}\,.$$

Altogether then, the first, third, and fifth of the six terms that will eventually make up the expression for $\widetilde{\Gamma}^i_{jk}$ amount to

$$\frac{\partial y^i}{\partial x^p}\,\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^r}{\partial y^k}\left(\frac{1}{2}g^{ps}\Big(\frac{\partial g_{qs}}{\partial x^r} + \frac{\partial g_{rs}}{\partial x^q} - \frac{\partial g_{qr}}{\partial x^s}\Big)\right).$$

By definition, what we have as the sum of these three terms is

$$\Gamma^p_{qr}\,\frac{\partial y^i}{\partial x^p}\,\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^r}{\partial y^k}\,.$$

In other words, this much of the expression would be precisely $\widetilde{\Gamma}^i_{jk}$ if we were dealing with a tensor of type $(1,2)$. The three remaining "subterms" that we must now discuss are the reason why it isn't a tensor.

This "discrepancy" amounts to

$$\frac{1}{2}g^{ps}g_{qr}\,\frac{\partial y^i}{\partial x^p}\,\frac{\partial y^l}{\partial x^s}\left(\frac{\partial}{\partial y^k}\Big(\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^r}{\partial y^l}\Big) + \frac{\partial}{\partial y^j}\Big(\frac{\partial x^q}{\partial y^l}\,\frac{\partial x^r}{\partial y^k}\Big) - \frac{\partial}{\partial y^l}\Big(\frac{\partial x^q}{\partial y^j}\,\frac{\partial x^r}{\partial y^k}\Big)\right).$$

When the derivatives of the products are expanded, we get six terms, four of which cancel out, and the portion of the expression inside the outer parentheses is

$$\frac{\partial x^r}{\partial y^l}\,\frac{\partial^2 x^q}{\partial y^k\,\partial y^j} + \frac{\partial x^q}{\partial y^l}\,\frac{\partial^2 x^r}{\partial y^j\,\partial y^k}\,.$$

Once again, the multiplication by the factor $\partial y^l/\partial x^s$ leads to $r = s$ in the first term and $q = s$ in the second, so that the full expression becomes

$$\begin{aligned} & \frac{1}{2}\frac{\partial y^i}{\partial x^p}\left(g^{ps}\left(q_{qs}\frac{\partial^2 x^q}{\partial y^j\,\partial y^k}+g_{rs}\frac{\partial^2 x^r}{\partial y^j\,\partial y^k}\right)\right)\\ = & \frac{1}{2}\frac{\partial y^i}{\partial x^p}\left(\delta^p_q\frac{\partial^2 x^q}{\partial y^j\,\partial y^k}+\delta^p_r\frac{\partial^2 x^r}{\partial y^j\,\partial y^k}\right)\\ = & \frac{\partial y^i}{\partial x^p}\,\frac{\partial^2 x^p}{\partial y^j\,\partial y^k}\,. \end{aligned}$$

This term is the perturbation that prevents the Christoffel symbols from forming a tensor. □

**Remark 6.4.** One may well ask why it matters that the Christoffel symbols are not the components of a tensor. After all, we have the chain rule as our guide, and we have just exhibited the formula for translating Christoffel symbols between different parametrizations. Since we know how to do the conversion, why do we care if these coefficients don't form a tensor?

A simple answer is that we are thereby warned off from a naive and incorrect conversion of these very important quantities when we change parameters. More important, it is precisely the "discrepancy" term when coordinates are changed that assures us that the covariant derivative *is* a tensor, as we shall now see.

**3.6. The covariant derivative.** Although the Christoffel symbols are not themselves the components of a tensor, the particular way they transform, just demonstrated, has the effect of making the covariant derivative a tensor. As shown in Chapter 6, that result is fundamental in proving that the Riemann curvature tensor is indeed a tensor. Unfortunately, the proof of this fact is even more messy and tedious than the demonstration of the transformation law for the Christoffel symbols. Again, we shall break the proof into manageable pieces. it really requires only one small combinatorial trick to make it work, which we shall introduce at the proper point. We begin by proving that the covariant derivative of one tangent vector (tensor of type $(1,0)$) with respect to another is a tensor of type $(1,0)$. This assertion means that if $\tilde{u}^j = u^q(\partial y^j/\partial x^q)$ and $\tilde{v}^k = v^r(\partial y^k/\partial x^r)$ when coordinates are changed from $x$ to $y$, and $\nabla_{\boldsymbol{u}}\boldsymbol{v} = w^l(\partial/\partial x^l)$, then $\tilde{w}^l = w^s(\partial y^l/\partial x^s)$. (The explicit formula for $w^l$ is shown below in the statement of the theorem.) Again, this is a matter of showing that if the covariant derivative of $\boldsymbol{v}$ with respect to $\boldsymbol{u}$ is written in $y$-coordinates, the result will be the same as taking the covariant derivative of $\boldsymbol{v}$ with respect to $\boldsymbol{u}$ in $x$-coordinates, then converting the $x$-coordinates of the resulting tangent vector to $y$-coordinates. Putting it another way, we need to show that the effect of the resulting tangent vectors on any $C^\infty$ function will be the same.

**Theorem 6.3.** *The covariant derivative $\nabla_{\boldsymbol{u}}\boldsymbol{v}$ of a tangent vector $\boldsymbol{v}$ with respect to a tangent vector $\boldsymbol{u}$ is a tensor of type $(1,0)$, that is,*

$$\left(\tilde{u}^j\frac{\partial\tilde{v}^i}{\partial y^j}+\tilde{u}^j\tilde{v}^k\widetilde{\Gamma}^i_{jk}\right)\frac{\partial}{\partial y^i}=\left(u^q\frac{\partial v^p}{\partial x^q}+u^q v^r\Gamma^p_{qr}\right)\frac{\partial}{\partial x^p}\,. \tag{17}$$

*In terms of components,*

$$\tilde{u}^j\frac{\partial\tilde{v}^i}{\partial y^j}+\tilde{u}^j\tilde{v}^k\widetilde{\Gamma}^i_{jk}=\frac{\partial y^i}{\partial x^p}\left(u^q\frac{\partial v^p}{\partial x^q}+u^q v^r\Gamma^p_{qr}\right).$$

PROOF. Just as we found when converting the Christoffel symbols, the conversion of the covariant derivative involves terms containing a product of six partial derivatives. Fortunately, when the summation is carried out, some of these factors mutually annihilate themselves, leaving much more manageable expressions.

We work backwards from the expression of the covariant derivative in $y$-coordinates to the corresponding expression in $x$-coordinates. Here is the covariant derivative in $y$-coordinates:

$$\nabla_{\boldsymbol{u}}\boldsymbol{v} = \Big(\tilde{u}^j \frac{\partial \tilde{v}^i}{\partial y^j} + \tilde{u}^j \tilde{v}^k \widetilde{\Gamma}^i_{jk}\Big)\frac{\partial}{\partial y^i}\,.$$

We now set about systematically replacing the tilde-topped terms with their expressions in $x$-coordinates. That is, we make the following substitutions:

$$\begin{aligned}
\frac{\partial}{\partial y^i} &\mapsto \frac{\partial x^p}{\partial y^i}\frac{\partial}{\partial x^p}\,, \quad \tilde{u}^j \mapsto u^s \frac{\partial y^j}{\partial x^s}\,, \quad \tilde{v}^k \mapsto v^t \frac{\partial y^k}{\partial x^t} \\
\frac{\partial \tilde{v}^i}{\partial y^j} &\mapsto \frac{\partial \tilde{v}^i}{\partial x^q}\frac{\partial x^q}{\partial y^j} = \frac{\partial v^t}{\partial x^q}\frac{\partial y^i}{\partial x^t}\frac{\partial x^q}{\partial y^j} + v^t \frac{\partial^2 y^i}{\partial x^q\,\partial x^t}\frac{\partial x^q}{\partial y^j}\,.
\end{aligned}$$

The last of these substitutions causes the first term

$$\tilde{u}^j \frac{\partial \tilde{v}^i}{\partial y^j}\frac{\partial}{\partial y^i}$$

to be expressed as a sum of two terms. The first of these terms is

$$u^s \frac{\partial y^j}{\partial x^s}\frac{\partial v^t}{\partial x^q}\frac{\partial y^i}{\partial x^t}\frac{\partial x^q}{\partial y^j}\frac{\partial x^p}{\partial y^i}\frac{\partial}{\partial x^p}\,.$$

As with the Christoffel symbols, the summations on $i$ and $j$ collapse to give the product $\delta^p_t \delta^q_s$; and so this expression gives a term

$$\delta^p_t \delta^q_s u^s \frac{\partial v^t}{\partial x^q}\frac{\partial}{\partial x^p} = u^q \frac{\partial v^p}{\partial x^q}\frac{\partial}{\partial x^p}\,.$$

This expression is precisely what we are hoping to see as part of the final expression, but we do have once again a "discrepancy" to account for, namely the terms involving second-order partial derivatives. That discrepancy amounts to

$$u^s \frac{\partial y^j}{\partial x^s} v^t \frac{\partial^2 y^i}{\partial x^q\,\partial x^t}\frac{\partial x^q}{\partial y^j}\frac{\partial x^p}{\partial y^i}\frac{\partial}{\partial x^p}\,.$$

The summation on $j$ once again provides a factor of $\delta^q_s$, and this discrepancy is seen to amount to

$$u^q v^t \frac{\partial^2 y^i}{\partial x^q\,\partial x^t}\frac{\partial x^p}{\partial y^i}\frac{\partial}{\partial x^p}.$$

At this point, we introduce a combinatorial device. Starting from the general relation

$$\frac{\partial y^i}{\partial x^q}\frac{\partial x^p}{\partial y^i} = \delta^p_q = \text{const.}\,,$$

and differentiating with respect to $x^t$, we find that

$$\begin{aligned}
0 &= \frac{\partial^2 y^i}{\partial x^q\,\partial x^t}\frac{\partial x^p}{\partial y^i} + \frac{\partial y^i}{\partial x^q}\frac{\partial}{\partial x^t}\Big(\frac{\partial x^p}{\partial y^i}\Big) \\
&= \frac{\partial^2 y^i}{\partial x^q\,\partial x^t}\frac{\partial x^p}{\partial y^i} + \frac{\partial y^i}{\partial x^q}\frac{\partial y^m}{\partial x^t}\frac{\partial^2 x^p}{\partial y^i\,\partial y^m}\,.
\end{aligned}$$

As a result, we find that the discrepancy is

$$- u^q v^t \frac{\partial y^i}{\partial x^q} \frac{\partial y^m}{\partial x^t} \frac{\partial^2 x^p}{\partial y^i \, \partial y^m} \frac{\partial}{\partial x^p}.$$

We now need to convert the remainder of the expression for $\nabla_{\boldsymbol{u}} \boldsymbol{v}$ from $y$-coordinates to $x$-coordinates and combine it with what we have just found. That remainder is

$$\tilde{u}^j \tilde{v}^k \widetilde{\Gamma}^i_{jk} \frac{\partial}{\partial y^i} = u^s \frac{\partial y^j}{\partial x^s} v^t \frac{\partial y^k}{\partial x^t} \Big( \Gamma^l_{qr} \frac{\partial y^i}{\partial x^l} \frac{\partial x^q}{\partial y^j} \frac{\partial x^r}{\partial y^k} + \frac{\partial y^i}{\partial x^l} \frac{\partial^2 x^l}{\partial y^j \, \partial y^k} \Big) \frac{\partial x^m}{\partial y^i} \frac{\partial}{\partial x^m} \, .$$

The second term in this expression is

$$u^s v^t \frac{\partial y^j}{\partial x^s} \frac{\partial y^k}{\partial x^t} \frac{\partial y^i}{\partial x^l m} \frac{\partial x^m}{\partial y^i} \frac{\partial^2 x^l}{\partial y^j \, \partial y^k} \frac{\partial}{\partial x^p} \, .$$

The summation on $i$ furnishes a factor $\delta^m_l$, and this term becomes

$$u^s v^t \frac{\partial y^j}{\partial x^s} \frac{\partial y^k}{\partial x^t} \frac{\partial^2 x^l}{\partial y^j \, \partial y^k} \frac{\partial}{\partial x^l}.$$

For the sake of clarity, let us replace the dummy indices of summation as follows $j \mapsto i$, $k \mapsto m$, $l \mapsto p$, $s \mapsto q$. This expression then becomes

$$u^q v^t \frac{\partial y^i}{\partial x^q} \frac{\partial y^m}{\partial x^t} \frac{\partial^2 x^p}{\partial y^i \, \partial y^m} \frac{\partial}{\partial x^p}.$$

In other words, this term exactly cancels the discrepancy that we found above. As for the first term involving the Christoffel symbol, it is

$$u^s \frac{\partial y^j}{\partial x^s} v^t \frac{\partial y^k}{\partial x^t} \Gamma^l_{qr} \frac{\partial y^i}{\partial x^l} \frac{\partial x^q}{\partial y^j} \frac{\partial x^r}{\partial y^k} \frac{\partial x^p}{\partial y^i} \frac{\partial}{\partial x^m} \, .$$

Here, the summations on $i$, $j$, and $k$ provide respectively the factors $\delta^p_l$, $\delta^q_s$, and $\delta^r_t$, and we get

$$u^q v^r \Gamma^p_{qr} \frac{\partial}{\partial x^p} \, .$$

Combining the two terms that we have now obtained, we have Eq. (17). □

**Remark 6.5.** At each point, the covariant derivative $\nabla_{\boldsymbol{u}} \boldsymbol{v}$ is determined by the values of the coefficients of $\boldsymbol{u}$ *at* the point and the values of the coefficients of $\boldsymbol{v}$ *in a neighborhood of* the point. For that reason, it is often defined as an operation on a contravariant *vector field* by a contravariant *vector* whose result is a contravariant *vector*.

**Remark 6.6.** It may seem puzzling that $\nabla_{\boldsymbol{u}} \boldsymbol{v}$, which is a *contra*variant vector field formed from two other contravariant vector fields $\boldsymbol{u}$ and $\boldsymbol{v}$, is yet called the *co*variant derivative of $\boldsymbol{v}$ with respect to $\boldsymbol{u}$. Keep in mind that it is the *operation* $(\boldsymbol{u}, \boldsymbol{v}) \mapsto \nabla_{\boldsymbol{u}} \boldsymbol{v}$ that is covariant, being a bilinear mapping of a pair of contravariant vectors.

The invariant (tensor) nature of the covariant derivative is a fundamental pillar supporting the use of parameters to compute the Riemann and Ricci curvature tensors. It guarantees that they are tensors, and hence that each observer is free to use whatever coordinates are most convenient.

## 4. Problems

**Problem 6.1.** Verify that the two parametrizations of the hemisphere given in this chapter have the transition mappings indicated. That is, show that, for example, the second equation becomes the identity $v = v$, when the values of $x$ and $y$ given in the third and fourth equations are substituted into the second equation.

**Problem 6.2.** Verify the relations between the metric coefficients of $A$ and those of $B$ as given in the text.

**Problem 6.3.** Verify the relation between the two first fundamental forms of $A$ and $B$ as given in the text.

**Problem 6.4.** Illustrate the results of the last two problems by considering the two parametrizations of the hemisphere of radius $R$ given as an example. Find the length of the portion of a great circle joining the points $(0.64R, 0.60R, 0.48R)$ and $(-0.64R, 0.60R, 0.48R)$ and the area between this great circle and the equator.

**Problem 6.5.** Verify the relation

$$EG - F^2 = \left|\frac{\partial(x,y)}{\partial(u,v)}\right|^2 (PR - Q^2)\,.$$

**Problem 6.6.** Show that a linear combination of tensors of type $(k,l)$ (with addition and scalar multiplication defined componentwise) is a tensor of type $(k,l)$.

**Problem 6.7.** The *tensor product* of a tensor field $S$ of type $(k,l)$ with components $S^{j_1\dots j_l}_{i_1\dots i_k}$ and a tensor field $T$ of type $(p,q)$ with components $T^{n_1\dots n_q}_{m_1\dots m_p}$ is defined as the multilinear operator $S \otimes T$ whose effect on a set of $k+p$ vector fields $\boldsymbol{u}_a$ and $l+q$ covector fields $\boldsymbol{v}^b$ is defined by

$$S \otimes T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_{k+m}, \boldsymbol{v}^1, \dots, \boldsymbol{v}^{l+n}) = \\ = S^{j_1\dots j_l}_{i_1\dots i_k} T^{n_1\dots n_q}_{m_1\dots m_p} u^{i_1}_1 \dots u^{i_k}_k u^{m_1}_{k+1} \dots u^{m_p}_{k+p} v^1_{j_1} \dots v^l_{j_l} v^{l+1}_{n_1} \dots v^{l+q}_{n_q}\,.$$

Show that $S \otimes T$ is a tensor field of type $(k+p, l+q)$.

**Problem 6.8.** If $T$ is a tensor field of type $(k,l)$, where $k$ and $l$ are both positive—let its components be $T^{j_1\dots j_l}_{i_1\dots i_k}$—the *contraction* of $T$ on a pair of indices $i_r$, $j_s$ is defined as the multilinear operator whose components are $T^{j_1\dots j_{s-1}\, m\, j_{s+1}\dots j_l}_{i_1\dots i_{r-1}\, m\, i_{r+1}\dots i_k}$, where, of course, summation over $m$ is required. Show that the contraction is a tensor of type $(k-1, l-1)$.

**Problem 6.9.** Let $G = g_{ij}$ be the metric on a manifold and $T$ a tensor field of type $(k,l)$ where $l > 0$, with the usual components $T^{j_1\dots j_l}_{i_1\dots,i_k}$. As mentioned in the text above, we can "lower" one of the indices $j_r$ by first forming the tensor $G \otimes T$, then contracting it on the lower index $j$ and the upper index $j_r$. Show that the result is a tensor of type $(k+1, l-1)$. (The covariant Riemann curvature tensor $R_{ijkl}$ is the result of lowering the upper index of the Riemann curvature tensor $R^i_{jkl}$.)

**Problem 6.10.** Show that for two systems of parameters that are *linearly* related, the Christoffel symbols do transform like a tensor of type $(1,2)$.

# Subject Index

# Name Index